THE NIH LIBRARY
NATIONAL INSTITUTES OF HEALTH

THE NIH LIBRARY
NATIONAL INSTITUTES OF HEALTH

MONOGRAPHS
JOURNAL OF THE NATIONAL CANCER INSTITUTE

First National AIDS Malignancy Conference

1998
Number 23

Contents

First National AIDS Malignancy Conference Monograph: Introduction vii
Ellen Feigal

Overview of the Epidemiology of Immunodeficiency-Associated Cancers 1
Valerie Beral, Robert Newton

Current Perspectives on the Molecular Pathogenesis of Virus-Induced Cancers in Human Immunodeficiency Virus Infection and Acquired Immunodeficiency Syndrome 7
Elliott Kieff

Human Papillomavirus Infection and Anogenital Neoplasia in Human Immunodeficiency Virus-Positive Men and Women 15
Joel M. Palefsky

Acquired Immunodeficiency Syndrome-Related Cancers: the Community Perspective 21
Michael Marco

Association of Non-Acquired Immunodeficiency Syndrome-Defining Cancers With Human Immunodeficiency Virus Infection 23
Charles S. Rabkin

Papillomaviruses and Cervical Cancer: Pathogenesis and Vaccine Development 27
Douglas R. Lowy, John T. Schiller

Cancers in Human Immunodeficiency Virus-Infected Children 31
Brigitta U. Mueller

Hodgkin's Disease in the Setting of Human Immunodeficiency Virus Infection 37
Alexandra M. Levine

Management of Cervical Neoplasia in Human Immunodeficiency Virus-Infected Women 43
Mitchell Maiman

Human Herpesvirus Type 8 and Kaposi's Sarcoma 51
Robin A. Weiss, Denise Whitby, Simon Talbot, Paul Kellam, Chris Boshoff

Some Aspects of the Pathogenesis of HIV-1-Associated Kaposi's Sarcoma 55
Robert C. Gallo

Clinical Overview: Issues in Kaposi's Sarcoma Therapeutics 59
Susan E. Krown

Kaposi's Sarcoma-Associated Herpesvirus-Encoded Oncogenes and Oncogenesis 65
Patrick S. Moore, Yuan Chang

Human Herpesvirus 8—the First Human *Rhadinovirus* 73
Frank Neipel, Jens-Christian Albrecht, Bernhard Fleckenstein

Novel Organizational Features, Captured Cellular Genes, and Strain Variability Within the Genome of KSHV/HHV8 79
John Nicholas, Jian-Chao Zong, Donald J. Alcendor, Dolores M. Ciufo, Lynn J. Poole, Robert T. Sarisky, Chuang-Jiun Chiou, Xiaoqun Zhang, Xiaoyu Wan, Hong-Guang Guo, Marvin S. Reitz, Gary S. Hayward

Immunotherapy for Epstein-Barr Virus-Associated Cancers 89
Cliona M. Rooney, Marie A. Roskrow, Colton A. Smith, Malcolm K. Brenner, Helen E. Heslop

(Contents continued on back cover)

Keep this essential journal coming direct to your desk.

Become a subscriber, today.

www.oup.co.uk/jnci/

The *Journal of the National Cancer Institute* remains the Number 1 journal in the field (*ISI Journal Citation Reports*, October 1996), publishing the best original research papers in oncology from around the world. Internationally acclaimed as the source for the most up-to-date news and information from the rapidly changing fields of cancer research and treatment, the *Journal of the National Cancer Institute* is now available by subscription from Oxford University Press.

Each issue is packed with extensive news about important cancer-related science, policy, politics, and people. Plus, editorials and reviews by experts in the field, book reviews, and commentary on timely topics.

And there's more!

Immediately upon subscribing, you will become a member of the **NCI Information Associates Program,** which includes these outstanding benefits:

- One copy of each new Journal of the National Cancer Institute Monographs*
- 24-hour access to PDQ, the NCI's comprehensive cancer information database
- Electronic Bulletin Board System
- Quarterly Newsletter
- Trained Member Service Representatives

For more information about the Information Associates Program, call 1-800-624-7890 (+301-496-7600 outside the USA).

*Monographs are available to individuals only on request. Institutional rate shown below automatically includes monographs.

NATIONAL CANCER INSTITUTE
INFORMATION ASSOCIATES PROGRAM

I wish to subscribe to the ***Journal of the National Cancer Institute.***

☐ I enclose payment for the full 1998 volume year (checks payable to Oxford University Press). I understand that I will receive all issues from January 1998.

☐ Please bill me for the full 1998 volume year.

☐ Please bill me for the partial current year plus 1998. I understand that I will receive the current issue through December 1998.

Name (please print) _____

Address _____

_____

Zip/Post code _____ Country _____

Please charge my

☐ MasterCard ☐ Visa ☐ American Express ☐ Diners ☐ JCB card.

Card number

1998 Subscription Rates

Volume 90 (24 issues)

OXFORD JOURNALS

Individuals	
US & Canada	☐ **One year: US$132**
UK & Europe	☐ **One year: £127**
Elsewhere	☐ **One year: US$180**
Institutions	
US & Canada	☐ **One year: US$190**
UK & Europe	☐ **One year: £170**
Elsewhere	☐ **One year: US$240**

Single issues are available upon request. Air-expedited delivery fee is included for subscriptions outside North America. Individual subscriptions must be paid for by personal check or credit card and sent to private addresses only. Canadian residents add 7% GST. Reduced rates are available to nonprofit organizations; please contact Oxford University Press for more information.

To begin your subscription, call 1-800-852-7323 (USA & Canada).

Or call 919-677-0977. Fax: 919-677-1714.

Outside the Americas, ring +44 (0)1865 267907 or fax +44 (0)1865 267485.

In the Americas, please return to: Oxford University Press, Journals Marketing (f), 2001 Evans Road, Cary, NC 27513, USA.

Elsewhere, please return to: Oxford University Press, Journals Marketing (f), Great Clarendon St., Oxford, OX2 6DP, UK.

MONOGRAPHS

JOURNAL OF THE NATIONAL CANCER INSTITUTE

Number 23 *1998*
ISSN 0027-8874
ISBN 0-19-922371-8

LIBRARY
MAY 29 1998
National Institutes of Health

EDITORIAL BOARD

Barnett S. Kramer
Editor-in-Chief

J. Gordon McVie
European Editor

Eric J. Seifter
Book Reviews Editor

J. Paul Van Nevel
News Editor

Frederic J. Kaye
Douglas L. Weed
Reviews Editors

Martin L. Brown
Economics Editor

ASSOCIATE EDITORS

Susan G. Arbuck
Frank M. Balis
William J. Blot
Peter M. Blumberg
John D. Boice, Jr.
Louise A. Brinton
Bruce A. Chabner
Ross C. Donehower
Susan S. Ellenberg
Suzanne W. Fletcher
Michael A. Friedman
Patricia A. Ganz
John K. Gohagan
Frank J. Gonzalez
Michael M. Gottesman
Peter Greenwald
Donald E. Henson
Susan M. Hubbard
Colin R. Jefcoate
Frederic J. Kaye
Hynda K. Kleinman
Theodore S. Lawrence
Bernard Levin
W. Marston Linehan
Marc E. Lippman
Scott M. Lippman
Darrell T. Liu
Dan L. Longo
Reuben Lotan
Douglas R. Lowy
Susan G. Nayfield
David L. Nelson
Kenneth Olden
David G. Poplack
Ross L. Prentice
Alan S. Rabson
Edward A. Sausville
Robert H. Shoemaker
David Sidransky
Richard M. Simon
Michael B. Sporn
Maryalice Stetler-Stevenson
J. Paul Van Nevel
Douglas L. Weed
Noel S. Weiss

STATISTICAL EDITORS

Janet W. Andersen
Jacques Benichou
Donald A. Berry
Barry W. Brown
Bernard F. Cole
Susan S. Ellenberg
Scott S. Emerson
Eric Feuer
Edmund A. Gehan
Barry I. Graubard
Sylvan B. Green
Susan G. Groshen
Richard A. Olshen
Philip C. Prorok
Timothy R. Rebbeck
Philip S. Rosenberg
Harland N. Sather
Daniel J. Schaid
Richard M. Simon
Donald M. Stablein
Robert E. Tarone
Sholom Wacholder

EDITORIAL ADVISORY BOARD

Hans-Olov Adami
Harry B. Burke
Pelayo Correa
Ethan Dmitrovsky
Merrill J. Egorin
Soldano Ferrone
Isaiah J. Fidler
Edward L. Giovannucci
Jean L. Grem
Curtis C. Harris
Emily L. Harris
Kathy J. Helzlsouer
Maureen M. Henderson
Gloria H. Heppner
Allan Hildesheim
Waun Ki Hong
William J. Hoskins
Peter J. Houghton
David H. Johnson
V. Craig Jordan
Ilan R. Kirsch
Kenneth H. Kraemer
Alexandra M. Levine
Martha S. Linet
Henry T. Lynch
Frank L. Meyskens
Anthony B. Miller
Malcolm S. Mitchell
David L. Page
Kenneth J. Pienta
David Schottenfeld
Richard K. Severson
Jerry W. Shay
Sandra M. Swain
G. Marie Swanson
Mario Sznol
Raymond Taetle
Ian Tannock
Peter R. Twentyman

Richard D. Klausner
Director, National Cancer Institute

Susan Molloy Hubbard
Director, International Cancer Information Center

Julianne Chappell
Chief, Scientific Publications Branch; Executive Editor

EDITORIAL STAFF

Scientific Editor: Ellen Feigal, M.D.

Monograph Coordinator: Elizabeth Horowitz

Manuscript Editors: Elaine Price Beck
Joan O'Brien Rodriguez
Elizabeth Horowitz

Editorial Assistant: Jamia V. Capehart

MARKETING STAFF

Marketing Director: Jean Griffin Baum

Press Contact: Ellen Scott

EDITORIAL POLICY: Manuscripts from key conferences dealing with cancer and closely related research fields, or a related group of papers on specific subjects of importance to cancer research, are considered for publication, with the understanding that they have not been published previously and are submitted exclusively to the *Journal of the National Cancer Institute Monographs.* All material submitted for consideration will be subject to review, when appropriate, by at least one outside reviewer and one member of the Editorial Board of the *Journal of the National Cancer Institute.* Opinions expressed by the authors are not necessarily those of the publisher or the editors.

Proposals for monographs should be submitted to the Editor-in-Chief, *Journal of the National Cancer Institute,* National Cancer Institute, P.O. Box 31110, Bethesda, MD 20824-1110.

Journal of the National Cancer Institute Monographs are available on request to members of the National Cancer Institute Information Associates Program (one copy of each monograph per member). Monographs are also available from Oxford University Press, Journals Subscription Department, 2001 Evans Road, Cary, NC 27513 USA. Telephone toll-free in the U.S. and Canada, 1-800-852-7323, or 919-677-0977. Fax: 919-677-1714. E-mail: jnlorders@oup-usa.org.

© Oxford University Press

First National AIDS Malignancy Conference

Proceedings of a Conference
Held at the
National Institutes of Health
Bethesda, Maryland
April 28–30, 1997

Conference Sponsor

National Cancer Institute
Richard Klausner, M.D.
Director

Contents

First National AIDS Malignancy Conference Monograph: Introduction vii
Ellen Feigal

Overview of the Epidemiology of Immunodeficiency-Associated Cancers 1
Valerie Beral, Robert Newton

Current Perspectives on the Molecular Pathogenesis of Virus-Induced Cancers in Human Immunodeficiency Virus Infection and Acquired Immunodeficiency Syndrome 7
Elliott Kieff

Human Papillomavirus Infection and Anogenital Neoplasia in Human Immunodeficiency Virus-Positive Men and Women 15
Joel M. Palefsky

Acquired Immunodeficiency Syndrome-Related Cancers: the Community Perspective 21
Michael Marco

Association of Non-Acquired Immunodeficiency Syndrome-Defining Cancers With Human Immunodeficiency Virus Infection 23
Charles S. Rabkin

Papillomaviruses and Cervical Cancer: Pathogenesis and Vaccine Development 27
Douglas R. Lowy, John T. Schiller

Cancers in Human Immunodeficiency Virus-Infected Children 31
Brigitta U. Mueller

Hodgkin's Disease in the Setting of Human Immunodeficiency Virus Infection 37
Alexandra M. Levine

Management of Cervical Neoplasia in Human Immunodeficiency Virus-Infected Women 43
Mitchell Maiman

Human Herpesvirus Type 8 and Kaposi's Sarcoma 51
Robin A. Weiss, Denise Whitby, Simon Talbot, Paul Kellam, Chris Boshoff

Some Aspects of the Pathogenesis of HIV-1-Associated Kaposi's Sarcoma 55
Robert C. Gallo

Clinical Overview: Issues in Kaposi's Sarcoma Therapeutics 59
Susan E. Krown

Kaposi's Sarcoma-Associated Herpesvirus-Encoded Oncogenes and Oncogenesis 65
Patrick S. Moore, Yuan Chang

Human Herpesvirus 8—the First Human *Rhadinovirus* 73
Frank Neipel, Jens-Christian Albrecht, Bernhard Fleckenstein

Novel Organizational Features, Captured Cellular Genes, and Strain Variability Within the Genome of KSHV/HHV8 79
John Nicholas, Jian-Chao Zong, Donald J. Alcendor, Dolores M. Ciufo, Lynn J. Poole, Robert T. Sarisky, Chuang-Jiun Chiou, Xiaoqun Zhang, Xiaoyu Wan, Hong-Guang Guo, Marvin S. Reitz, Gary S. Hayward

Immunotherapy for Epstein-Barr Virus-Associated Cancers 89
Cliona M. Rooney, Marie A. Roskrow, Colton A. Smith, Malcolm K. Brenner, Helen E. Heslop

Genetic Basis of Acquired Immunodeficiency Syndrome-Related Lymphomagenesis 95
Gianluca Gaidano, Antonino Carbone, Riccardo Dalla-Favera

Clinical Management of Human Immunodeficiency Virus-Associated Non-Hodgkin's Lymphoma 101
Lawrence D. Kaplan

First National AIDS Malignancy Conference Monograph: Introduction

*Ellen Feigal**

Scientists, physicians, health care workers, and community and patient advocates from around the world gathered at the first national forum focused on acquired immunodeficiency syndrome (AIDS)-associated cancers in April 1997 on the grounds of the National Institutes of Health campus in Bethesda, MD. The first National AIDS Malignancy Conference, sponsored by the National Cancer Institute, featured multidisciplinary presentations on the epidemiology, biology, virology, immunology, and treatment of cancers in the setting of human immunodeficiency virus (HIV) infection. Cancers have been associated with HIV since the beginning of the epidemic in the late 1970s. With almost 1 million individuals infected with HIV across the United States and 10 million infected worldwide, AIDS-associated cancers offer a unique perspective on the interplay of viruses, immune dysregulation, and cancer pathogenesis. This monograph is a comprehensive compilation of papers from 18 plenary speakers.

Overview of Cancers, Viruses, and Immunodeficiency

Dr. Valerie Beral opened the conference with a presentation of the epidemiologic evidence on the effect of immunosuppression on cancer risk. Immunodeficiency, whether congenital, therapeutic, or infectious in origin, increases the risk of certain, but not all, types of cancer. A common feature of these cancers is that specific infectious agents appear to be important in their etiology. The study of cancer in immunodeficient populations offers a unique opportunity to investigate the role of the immune system in controlling the development, growth, and dissemination of tumors. Such studies have already contributed substantially to knowledge about the role of infectious agents in human cancer.

Given the strategic role that infectious agents may have in the etiology of these cancers, Dr. Elliott Kieff discussed the molecular mechanisms by which human papillomavirus (HPV), Epstein-Barr virus (EBV), human herpesvirus type 8 (HHV-8), and HIV persist and effect changes in cell growth that result in malignancy. Pharmacologic strategies to inhibit the mechanisms by which these viruses cause persistent infection and cell growth transformation may be useful in preventing and treating these cancers. Those approaches may also be useful in considering prevention and treatment of the underlying HIV and other opportunistic infections.

The focus of the overview talks next centered on a discussion of the epidemiologic role of HPV in the development of anogenital neoplasia in both men and women, the higher incidence of HPV in the HIV-infected population, and clinical features of the disease in the HIV-infected as contrasted to the immunocompetent individual. Dr. Joel M. Palefsky postulated that the higher incidence of HPV may reflect loss of systemic immune response to HPV antigens or local HPV–HIV interactions at the tissue or cellular level.

Activist Community

The AIDS activist community is a source of expertise on how science and clinical medicine interact with politics and policy. Activists have had an enormous impact on how the research community shapes its directions, and they have opened the doors for other community and patient advocates to be involved in the decision-making process that ultimately has an impact on patients and individuals at risk for the disease. In his plenary talk, Mr. Michael Marco challenged the National Cancer Institute, the Food and Drug Administration, and the biotechnology and pharmaceutical industries to become more engaged in the basic and clinical scientific communities. A primary emphasis of his presentation was asking how the National Cancer Institute can assure that patients with AIDS and cancer will receive the best medical care for their cancer as well as for their underlying HIV infection.

Non-AIDS-Defining, HPV-Associated, and Pediatric Cancers

A frequently asked question is whether new cancers are emerging in the HIV-infected population. Dr. Charles S. Rabkin examined evidence on whether there are age-related as well as geographic variations in their prevalence. Varying levels of evidence link several additional neoplasms to HIV infection, including Hodgkin's disease and anogenital intraepithelial neoplasia, although invasive disease is still uncertain for both cervical and anal cancers. Leiomyosarcoma and benign leiomyomas are increased in HIV-infected children, but they are unusual in HIV-infected adults. Conjunctival cancer is seen in HIV-infected individuals from sub-Saharan Africa, but it is uncommon in Western countries.

The strong epidemiologic data linking particular subsets of HPV to anogenital disease, combined with knowledge gained from molecular biology and immunology, have spurred recent efforts that have focused on the development of vaccines against HPV. Dr. Douglas R. Lowy raised the possibility that virus-like particles, resembling native HPV capsids, might be effective immunogens for a prophylactic HPV vaccine. Dr. Mitchell Maiman provided his insight on the diagnosis and management

**Affiliation of author:* Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD.

Correspondence to: Ellen Feigal, M.D., National Institutes of Health, Bldg. 31, Rm. 3A44, Bethesda, MD 20892.

of cervical intraepithelial neoplasia and cervical cancer in HIV-infected women.

Children with HIV infection have a lower incidence of cancer than adults with HIV infection, but they appear to have unusual and extremely rare tumors, such as leiomyosarcomas and leiomyomas, and a high prevalence of lymphoproliferative disorders. Dr. Brigitta U. Mueller noted that there are approximately 130 cases of cancer per million non-HIV-infected children (0.013%) per year. A conservative estimate is that children with HIV infection appear to have at least a 100-fold higher incidence of cancers. The need for oncologic and infectious disease expertise in treating these types of uncommon tumors was stressed.

Hodgkin's disease represents another cancer that is not AIDS-defining, although its clinical and pathologic characteristics differ from those in the immunocompetent setting. Dr. Alexandra M. Levine compared and contrasted the characteristics of Hodgkin's disease in the immunocompetent and immunodeficient settings.

Kaposi's Sarcoma (KS): KSHV/HHV-8, Inflammatory Cytokines, and Treatment

In the next series of overviews, Dr. Robin A. Weiss set the scene with a presentation on the epidemiology, current serologic tests, and various KS-associated herpesvirus (KSHV)/HHV-8-encoded proteins that may play a role in the promotion of cellular growth. Later in the conference, molecular biologists analyzed the viral genome and described how it replicates and functions.

Challenging prevailing views on KSHV/HHV-8 and its etiologic relationship to disease, Dr. Robert C. Gallo suggested that the role of HIV-1 is more active in the KS disease process than simply promoting immunodeficiency. He suggested that HIV-1 acts directly by promoting an increase in inflammatory cytokines, which, through sustained release, influences early stage KS by inciting local micro-inflammatory responses and affects growth of the inflammatory cells through the Tat protein.

The clinical aspects of KS were presented by Dr. Susan E. Krown. Four questions critical to the development of improved therapeutic and prophylactic strategies included the following: 1) Can we identify risk factors and predict who will develop KS? 2) Can we translate hypotheses about pathogenesis into improved therapeutic or prophylactic strategies, e.g., target new blood vessel development or inhibit inflammatory cytokines? 3) How does improved anti-HIV therapy have an impact on KS treatment? 4) How can we best evaluate the benefit of therapy beyond conventional measures of tumor response?

KSHV/HHV-8 Symposium

The KSHV/HHV-8 symposium was led by Dr. Patrick S. Moore, co-discoverer with Dr. Yuan Chang and colleagues of KSHV/HHV-8. Molecular biologic studies of KSHV have identified a number of potential oncogenes that may contribute to neoplasia. With the publication of the full-length genomic sequence, new opportunities to investigate its molecular biology and virology are becoming available. This presentation provided a road map for subsequent papers in this symposium on the major features of the KSHV genome and its potential oncogenes. The virus is already providing unique insights into tumorigenesis and may serve as an important model for virus-induced oncogenesis.

The first known human member of the genus *Rhadinovirus* is HHV-8/KSHV. Dr. Bernhard Fleckenstein noted that acquisition of genes from the host cell genome is a common feature of most herpesviruses and of rhadinoviruses in particular. Although different rhadinoviruses acquire different host-cell genes, their putative functions apparently converge to achieve three common goals: 1) enhance DNA replication independent from the cell cycle, 2) expand the pool of infectable cells, and 3) counteract the host's responses to infection.

Dr. Gary S. Hayward delved into the HHV-8 genome structure and strain differences. These differences would provide the opportunity to address the origins, geographical preferences, transmissibility, and disease association.

Lymphomas: EBV, Molecular Pathogenesis, and Treatment

EBV-associated lymphoproliferative disease is a frequently fatal complication of organ transplantation and HIV infection. Dr. Cliona M. Rooney presented her study on the safety and efficacy of adoptively transferred, gene-marked, virus-specific cytotoxic T lymphocytes as prophylaxis and treatment of EBV-associated lymphoproliferative disease and its implications for HIV-related cancers.

Molecular pathogenesis of AIDS-associated non-Hodgkin's lymphoma is a complex process involving both host factors and the accumulation of genetic lesions within the tumor clone. Dr. Riccardo Dalla-Favera reviewed the pattern of molecular lesions in these tumors and several distinct pathogenic pathways in AIDS-related lymphomagenesis. These pathways selectively associate with the different clinicopathologic variants of AIDS-associated non-Hodgkin's lymphoma. On the basis of the differences in molecular characteristics and presumed mechanisms of pathogenesis among the AIDS-associated non-Hodgkin's lymphomas and given the relatively modest survival of patients treated with conventional cytotoxic chemotherapy, Dr. Lawrence D. Kaplan suggested that development of future therapeutic approaches should take advantage of some of these unique molecular characteristics.

This monograph is just a taste of the energizing atmosphere of the first National AIDS Malignancy Conference. Multidisciplinary, interactive discussions took place during the abstract presentations and poster sessions. The Second National AIDS Malignancy Conference occurred on April 6–8, 1998. Next-day summaries for those who were unable to attend the conference are available on http://www.healthcg.com.

I would like to acknowledge and thank Dr. Richard D. Klausner and Dr. Robert Wittes for their guidance and support and Dr. James Goedert and the other members of the 16-person program steering committee for their time and effort in the organization of this conference.

Overview of the Epidemiology of Immunodeficiency-Associated Cancers

*Valerie Beral, Robert Newton**

Immunodeficiency, be it congenital, therapeutic, or infectious in origin, increases the risk of certain, but not all, types of cancer. A common feature of these cancers is that specific infectious agents appear to be important in their etiology, not only in immunodeficient subjects but also in the general population. People with acquired immunodeficiency syndrome (AIDS) are at an increased risk of Kaposi's sarcoma, non-Hodgkin's lymphoma, Hodgkin's disease, squamous cell carcinoma of the conjunctiva, and childhood leiomyosarcoma. It is striking that most of these cancers have been associated with specific human herpesvirus (HHV) infections: HHV-8 with Kaposi's sarcoma and the closely related Epstein-Barr virus with non-Hodgkin's lymphoma, Hodgkin's disease, and possibly also with childhood leiomyosarcoma. Moreover, similar associations between these viruses and cancer have been found, albeit inconsistently, in people who are not immunosuppressed. Further research is needed to establish whether the risk of other cancers is also increased in people with AIDS, although, if so, the cancers are likely to be rare or to have comparatively small associated relative risks. Existing evidence suggests that there may be no marked increase in the risk of two common cancers that are known to be caused by infectious agents—hepatocellular carcinoma and invasive carcinoma of the uterine cervix. The apparent lack of an increase in invasive cervical cancer is unexpected and needs further investigation, especially since the prevalence of cervical infection with human papillomaviruses and of low-grade preneoplastic changes in the cervical epithelium is increased in women with AIDS. With the prospect of improved survival in people with AIDS, the effect of immunosuppression on cancer is likely to become an increasingly important issue. [Monogr Natl Cancer Inst 1998;23:1–6]

The study of cancer in immunodeficient populations offers a unique opportunity to investigate the role of the immune system in controlling the development, growth, and dissemination of tumors. Such studies have already contributed substantially to knowledge about the role of infectious agents in human cancer. This article reviews the epidemiologic evidence about the effect of immunosuppression on cancer risk.

Immunodeficiency and Cancer: The Evidence Before the Acquired Immunodeficiency Syndrome (AIDS) Epidemic

It has long been thought that the immune system plays a vital role in the etiology of cancer. In 1965, Nobel laureate Sir McFarlane Burnet argued that immunosurveillance was a central mechanism by which tumor development was kept in check and predicted that individuals who were immunosuppressed would be at an increased risk of cancer *(1)*. It was already known that children with rare congenital defects of their immune system, such as X-linked gammaglobulinemia or ataxia telangiectasia, were at increased risk of lymphoma, but the number of children with such defects was exceedingly small and so it was not possible to tell whether they were also at an increased risk of other types of cancer *(2)*. Since the 1970s, with the increasing use of immunosuppressive drugs in relation to tissue transplantation, it has been possible to investigate Burnet's hypothesis in detail. Studies of individuals on long-term immunosuppressive drug therapy have shown that such people were at an increased risk of certain, but not all, types of cancer. The most marked increases were for non-Hodgkin's lymphoma (the most commonly reported tumor in most studies of immunosuppressed transplant recipients), Kaposi's sarcoma, hepatocellular carcinoma, and squamous cell carcinoma of the skin, including lip and vulval cancers *(3,4)*. The magnitude of the increase in the relative risk of these tumors was very large indeed. For example, there was about a 100-fold increase in the risk of Kaposi's sarcoma and a 10-fold or greater increase in the risk of other cancers. Some studies *(4)* reported an increase in Hodgkin's disease, and others *(3)* reported an increase in cervical cancer, but it is not clear whether this was for premalignant epithelial changes in the cervix and/or for invasive cervical cancer *(3)*.

The findings in transplant recipients suggested that immunosuppression led to the selective development of cancers that were caused by infections. The first specific infectious agent implicated as causing cancer in transplant recipients was the Epstein-Barr virus (EBV), which has been consistently identified in transplant-related lymphomas *(5)*. Human herpesvirus (HHV)-8 has been associated with Kaposi's sarcoma, and hepatitis viruses B and C (HBV and HCV, respectively) have been associated with most hepatocellular carcinomas *(6)*. In addition, human papillomavirus (HPV) types 16/18 have been associated with cervical cancer, and HPV types 5/8 appear to be responsible for some skin cancers in the immunosuppressed *(7)*.

The clinical course of the cancers that occur in immunosuppressed transplant recipients tends to be more aggressive than in the general population. It has also been noted that the cessation of immunosuppressive therapy can halt or even reverse tumor

**Affiliation of authors:* Imperial Cancer Research Fund, Cancer Epidemiology Unit, Gibson Building, Radcliffe Infirmary, Oxford, U.K.

Correspondence to: Valerie Beral, M.D., ICRF Cancer Epidemiology Unit, Gibson Bldg., Radcliffe Infirmary, Oxford OX2 6HE, U.K.

See "Note" following "References."

© Oxford University Press

growth *(8)*. Furthermore, similar risk factors seemed to determine who developed the cancer, irrespective of immune status. For example, the transplant recipients who developed skin cancer tended to be fair skinned and have excessive lifetime exposures to the sun *(9)*.

By the late 1970s, before the AIDS epidemic, it was widely believed that immunosuppression led to the selective development of cancers that were caused by infectious agents. However, because the immunodeficient patients tended to be ill for other reasons, it was not always clear whether the cancers that occurred among them were due to the immunodeficiency itself or to the underlying medical conditions, related exposures, or even to the immunosuppressive drugs themselves. For example, the increased risk of hepatocellular carcinoma seen in transplant recipients may have been due to the high prevalence of HBV or HCV infection in this group, and the high rates of cervical cancer may have been due to more frequent screening for that cancer in transplant recipients than in the general population.

Cancers Associated With AIDS

The AIDS epidemic has provided an unprecedented opportunity to study the effects of immunosuppression on cancer. The fact that large numbers of people throughout the world are infected with the human immunodeficiency virus (HIV) permits the study of the effects of immunosuppression on cancer risk on a scale and in populations that had not been studied before. Furthermore, the reason for the immunodeficiency in people with AIDS was not the same as that for the populations studied previously. As will be discussed below, many of the increases in cancer risk found in people with AIDS are similar to the findings in immunodeficient children and in transplant recipients, suggesting that it is the impairment of immune function, rather than other factors, that is leading to the appearance of these tumors.

Cancers Definitely Increased in People With AIDS

Although many cancers have been reported to be increased in people with AIDS, for only five cancers is the evidence sufficiently strong and consistent that it is possible to conclude that there is a definite increase in risk. These cancers are listed in Table 1. The evidence for three of those five cancers—Kaposi's sarcoma, non-Hodgkin's lymphoma, and squamous cell carcinoma of the conjunctiva—is well established and has been reviewed in depth elsewhere *(6,10)*. Recent evidence for the other two cancers—Hodgkin's disease and childhood leiomyosarcoma—also suggests a definite increase in risk associated with HIV infection *(11–13)*. The relative risks for HIV-seropositive compared with HIV-seronegative people for the five cancers listed in Table 1 tend to be very large—generally 10-fold or greater. Furthermore, each of those cancers is believed to be caused by a specific infectious agent: HHV-8 for Kaposi's sarcoma, HPV 16/18 for conjunctival cancers (although the evidence for this is not consistent), and EBV for non-Hodgkin's lymphoma and Hodgkin's disease and possibly, also, for leiomyosarcoma in children.

Table 1. Cancers that are definitely increased among human immunodeficiency virus (HIV)-infected people

Cancer	Approximate relative risk in HIV-seropositive individuals	Infectious agent*
Kaposi's sarcoma	10 000	HHV-8
Non-Hodgkin's lymphoma	50	EBV
Hodgkin's disease	10	EBV
Squamous cell carcinoma of the conjunctiva	10	Possibly HPV 16/18
Childhood leiomyosarcoma	Unclear	Possibly EBV

*HHV = human herpesvirus; EBV = Epstein-Barr virus; HPV = human papillomavirus.

Some, but not all, of the cancers listed in Table 1 have been associated with other forms of immunodeficiency. Kaposi's sarcoma and non-Hodgkin's lymphoma have been strongly associated with both immunosuppressive drug therapy and with AIDS. Hodgkin's disease has also been associated with both types of immunosuppression, although the relative risks are not as large as those for the other two cancers *(4)*. Some of the associations found in people with AIDS that have not been reported in transplant recipients can probably be explained by the fact that HIV infection affects a much broader range of people living in different parts of the world. For example, squamous cell carcinoma of the conjunctiva is common in equatorial Africa, where HIV infection is common, but where tissue transplantation is rare. Also, leiomyosarcoma appears to occur exclusively as a rare complication of HIV infection in children, and few children have been given long-term immunosuppressive drug therapy.

Other Cancers in People With AIDS

The available evidence suggests that people with AIDS are not experiencing large increases in the risk of most types of cancer. Cancer trends for men living in San Francisco, where HIV prevalence is relatively high, show marked increases over time in the incidence of Kaposi's sarcoma and non-Hodgkin's lymphoma only *(14)*. Furthermore, in a case–control study of about 1000 people with all types of cancer in South Africa, the only cancers that were significantly increased in HIV-seropositive people compared with HIV-seronegative people were Kaposi's sarcoma and non-Hodgkin's lymphoma *(15)*. Results broadly similar to those from South Africa were found in a case–control study of 250 people in Rwanda *(16)*. Likewise, record linkage of data from AIDS registries and cancer registries in the United States and in Australia has tended to find increased risks mainly for the cancers listed in Table 1 *(11,17)*.

In the Australian study linking information from an AIDS registry to cancer registry data, Grulich et al. *(11)* reported a relative risk of 5.8 (95% confidence interval [CI] = 1.2–17) for multiple myeloma. There have been a number of case reports of plasmacytomas occurring in people with AIDS *(6)*. It seems likely, therefore, that some form of plasmacytoma may be a rare consequence of HIV-associated immunosuppression, although, at this stage the evidence is not as firm as it is for the five cancers listed in Table 1. Some researchers have linked multiple myeloma to infection with EBV, and others have linked it to infection with HHV-8 *(6,18)*.

Perhaps the most fascinating results are the consistent reports that people with AIDS do not appear to be at an increased risk for two common cancers that are known to be caused by infectious agents: invasive cervical cancer and hepatocellular cancer *(6)*.

Cervical Cancer

Cervical cancer is the most common cancer among women in Africa, and although HIV infection is highly prevalent in central and southern Africa, no epidemic of cervical cancer has been observed. By contrast, there are marked epidemics of AIDS-related Kaposi's sarcoma in these African countries *(19,20)*. Fig. 1 summarizes the relative risk of invasive cervical cancer in HIV-seropositive subjects compared with HIV-seronegative subjects from two published studies *(15,16)* and preliminary results from another study *(21)* conducted in three African countries, which together total 363 women with the tumor. The relative risk for invasive cervical cancer is not elevated in any study. The overall relative risk is 0.8 (95% CI = 0.5–1.4), thus arguing against the possibility of a large increase in risk. It should be noted that the relative risk estimates are adjusted by age, but not by sexual practices, and the failure to adjust by sexual practices may even lead to a slight overestimate of the relative risk. Although some of the HIV-seropositive subjects may not be severely immunosuppressed, such a bias would act to reduce the relative risk, but not to eliminate an association altogether.

There are very few data about invasive cervical cancer in North America or Europe, chiefly because of the low prevalence of HIV infection in women in those countries, but also because women in the West tend to have regular Pap smears and so preclinical cervical neoplasia is generally treated early. However, the few studies in the West that have looked at this question have also failed to find an increase in the risk of invasive cervical cancer for HIV-infected women *(6,22)*.

In contrast to the apparent lack of an increase in invasive cervical cancer, HIV infection is associated with a definite increase in the expression of HPV by cervical cells and in the prevalence of apparently premalignant cervical lesions. Numerous studies have shown markedly higher prevalences of cervical HPV infection, by either single or multiple types, among HIV-infected women. There is also a corresponding increase in the prevalence of low-grade cervical lesions; however, few women with high-grade cervical lesions or with *in situ* cervical cancer have been found in these studies *(6)*.

At this stage, there is no obvious explanation as to why HIV infection appears to increase the risk of low-grade cervical lesions, but not the risk of invasive cervical cancer. It has been suggested that HIV-infected women in Africa may die of other causes before they have time to develop invasive cervical cancer. However, cervical cancer is so common in Africa that there should be many women who already had premalignant cervical lesions and even *in situ* cancer prior to HIV infection. If immunosuppression associated with HIV infection hastened the clinical course of disease in such women, it is surprising that an epidemic of cervical cancer has not been seen in Africa and that an increase in the risk of invasive cervical cancer has not been found in association with HIV infection.

The possibility that there is little or no increase in the risk of invasive cervical cancer among HIV-infected women is worth considering, and, if so, this observation offers an important clue to the pathogenesis of cervical cancer. In women who are chronically infected with HPV, the immune system may play an important role in determining what type of cervical lesion ultimately develops. Low-grade cervical lesions mainly reflect HPV infection, and only a small proportion of women with these lesions go on to develop invasive disease. The HPV infections that occur in association with HIV infection may have little immunologic control, and this possibility may favor the persistence of low-grade cervical lesions that do not progress. By contrast, the typical, chronic HPV infections in women who are not immunosuppressed and have strong immunologic control may favor progression to invasive disease.

Hepatocellular Carcinoma

It is curious that there appears to be no increase in the risk of hepatocellular carcinoma in people with AIDS. Fig. 2 summarizes the results from two case–control studies in Africa *(15,16)*, preliminary results from a case–control study in Uganda *(21)*, and from one cohort study of patients with hemophilia in the U.K. *(23)*. In no study is there an increase in the relative risk of hepatocellular carcinoma in HIV-seropositive subjects compared with HIV-seronegative subjects. The overall relative risk is 0.8 (95% CI = 0.5–1.5). Hepatocellular carcinoma is sometimes difficult to diagnose in the absence of modern techniques, and it might be argued that this absence contributes to the absence of an association with HIV infection in Africa. Nevertheless, the study of U.K. hemophiliacs found a 17-fold overall increase in the risk of liver cancer (because of the high prevalence of HBV and HCV in hemophiliacs), but no difference in

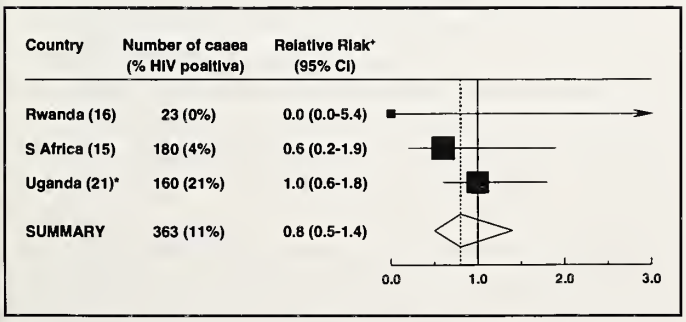


Fig. 1. Summary of results from studies investigating the risk of invasive cervical cancer in human immunodeficiency virus (HIV)-seropositive subjects compared with HIV-seronegative subjects. CI = confidence interval, * = preliminary results, and + = relative risk in HIV-seropositive subjects compared with HIV-seronegative subjects.

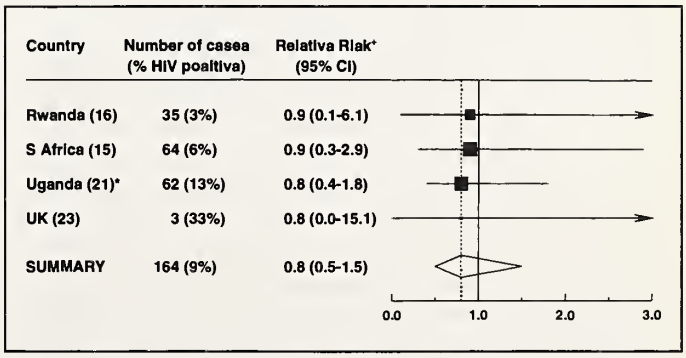


Fig. 2. Summary of results from studies investigating the risk of hepatocellular carcinoma in human immunodeficiency virus (HIV)-seropositive subjects compared with HIV-seronegative subjects. CI = confidence interval, * = preliminary results, and + = relative risk in HIV-seropositive subjects compared with HIV-seronegative subjects.

cancer risk between HIV-seropositive subjects and HIV-seronegative subjects, although numbers were small. Also, there was no evidence of a rise in hepatocellular cancer incidence in populations in the United States with a high prevalence of HIV infection *(6,14)*. Therefore, it is worth considering that, as with cervical cancer, the mechanism of hepatic carcinogenesis by HBV and HCV may be dependent on a strong immunologic response by the host or that it takes so long for the cancers to develop that people die of other causes.

Other Cancers

There may be other cancers whose incidence is increased in association with HIV infection but, if so, they are probably rare and the associated relative risks are not likely to be large. Increases in oral, testicular, skin, brain, lung, breast, and thyroid cancers have been suggested *(9)*. Different researchers have tended to report excesses of different types of cancer, and there was no consistent elevation in the relative risk for any particular tumor.

With the prospect of improved survival for HIV-infected people, it will become increasingly important to know more about the risk of cancer in these subjects. Further research is needed to clarify which other tumors are and which are not increased in people with AIDS and, if so, the magnitude of the associated relative risk. In particular, there is a need for further record linkage studies in populations where HIV prevalence is low and for further case–control studies in populations where HIV prevalence is high. In the meantime, because the numbers of other cancers reported in any single study tend to be small, it would be valuable to combine the results from existing studies.

Clinical and Biologic Features of the Cancers

Ever since the first cases of Kaposi's sarcoma were described in homosexual men, heralding the beginning of the HIV epidemic, it was clear that the clinical characteristics of the disease differed markedly from those of the usually indolent condition, generally affecting the skin of the lower limbs of elderly men, that had been described previously *(24)*. Although identical histologically, HIV-associated Kaposi's sarcoma often presents with multiple lesions affecting both the skin and internal organs, and survival drops from 10–15 years for HIV-seronegative individuals to about 14–18 months for HIV-seropositive individuals *(25,26)*. Such an aggressive presentation has occasionally been seen in HIV-seronegative children in Uganda *(27,28)*. Immunosuppression also affects the ease with which HHV-8 can be detected by polymerase chain reaction (PCR) in the peripheral blood of Kaposi's sarcoma patients. The frequency of detection via PCR increases, while the CD4 count declines *(29)*.

HIV-associated non-Hodgkin's lymphomas, although histologically heterogeneous, often exhibit pleomorphic features encompassing a range of overlapping subtypes (including Burkitt's lymphoma) and can be characterized by a number of features, including an aggressive clinical course. High-grade disease is common and extranodal sites are often involved, with lesions in the central nervous system being virtually unknown, except in the immunosuppressed. Similarly, Hodgkin's disease among HIV-infected individuals is clinically unusual, generally presenting at a late stage with extranodal dissemination common. The predominant histologic subtypes are mixed cellularity and lymphocyte-depleted, which are relatively rare in the HIV-uninfected population.

Whether these differences in the presentation and the clinical course between HIV-infected and uninfected individuals reflect differences in disease etiology or the host response is unclear, although the latter may be more likely. Certainly, infected individuals have a higher viral load of EBV when they are immunosuppressed and possibly also of HHV-8, although the evidence is less clear *(29,30)*.

Other Risk Factors

An individual's risk of developing a specific cancer is affected by various factors, including age, sex, and exposure to environmental factors. As yet, there is only limited information about the characteristics and exposures of HIV-infected people who develop cancer, but such people tend to have characteristics and exposures similar to those of HIV-uninfected people with a similar cancer.

The main determinant of Kaposi's sarcoma risk is infection with HHV-8, and this applies both to people with and without AIDS *(31)*. More is known about the personal characteristics of people with AIDS-related Kaposi's sarcoma than of people without AIDS. The main behavioral risk factor for AIDS-related Kaposi's sarcoma in western populations is sex between men *(32)*. This fact suggests that HHV-8 may be transmitted by sexual contact between men, and there is now some evidence that this is so *(33,34)*. Whether sex between men is also a risk factor for HIV-negative Kaposi's sarcoma is unknown. Certainly, Kaposi's sarcoma was far more common in men than in women before the AIDS epidemic; thus, it is possible that HHV-8 was transmitted sexually between men in the past, even though its prevalence in most countries was probably low. The recent spread of HIV in the population may have led to an increase in the spread of HHV-8 infection (because immunosuppression may facilitate transmission of HHV-8), resulting in an increase in Kaposi's sarcoma even in people who are not infected by HIV. For example, Kaposi's sarcoma is being increasingly reported in HIV-seronegative homosexual men in New York and in HIV-seronegative children in Africa *(6,27)*.

Sex between men is not the only behavioral risk factor for Kaposi's sarcoma, since the disease also occurs in women and in children. The incidence of Kaposi's sarcoma varied markedly across the world before the AIDS epidemic and was particularly frequent in Africa and infrequent in the U.K. (Table 2). The geographic variation in Kaposi's sarcoma before the AIDS epidemic reflects the geographic variation in the prevalence of HHV-8 infection and is similar to the geographic variation in the proportion of women with AIDS who also had Kaposi's sarcoma (Table 2) *(31,32,35–37)*. This variation suggests that the main determinant of Kaposi's sarcoma risk in different countries is the prevalence of HHV-8 infection.

Burkitt's lymphoma is associated with EBV infection both in people with and without AIDS. Burkitt's lymphoma is also far more common in males than in females and shows a characteristic peak at age 10–19 years, both in people with AIDS and in the general population *(38)*. EBV infection has also been associated with Hodgkin's disease, both in HIV-seropositive subjects and in HIV-seronegative subjects *(39,40)*.

Table 2. Geographic variation in the incidence of Kaposi's sarcoma before the acquired immunodeficiency syndrome (AIDS) epidemic, prevalence of human herpesvirus-8 (HHV-8) infection, and proportion of women with AIDS who have Kaposi's sarcoma*

Region or country	Kaposi's sarcoma incidence per million before the AIDS epidemic	Approximate prevalence of HHV-8 in human immunodeficiency virus-seronegative individuals, %	Proportion of women with AIDS who also have Kaposi's sarcoma, %
U.K.	0.1	0	0
United States	1.8	0	1
Italy	6.6	4	2
Uganda	50	50	10

**See* text for references.

For non-Burkitt's, non-Hodgkin's lymphomas, the proportion of people with AIDS who have these lymphomas increases with age, is greater in whites than in blacks, and is greater in men than in women, similar to the pattern of risk seen in the general population *(38)*. EBV has been found in about half the AIDS-related non-Burkitt's non-Hodgkin's lymphomas and in virtually all the AIDS-related lymphomas of the central nervous system *(5)*. EBV has also been found, although not consistently, in subjects with non-Burkitt's non-Hodgkin's lymphoma in the general population.

Although the evidence to date is not strong, it appears that cancers that occur in the immunosuppressed population have a similar etiology to the cancers that occur in the general population. Infectious agents are consistently found in a large proportion of tumors in the immunosuppressed subjects but generally in a smaller proportion of similar tumors in immunocompetent people. The question is whether the same infectious agents may be involved in the etiology of the same type of cancers both in the immunosuppressed and in the general population. Present diagnostic techniques are not sensitive enough to identify the infectious agents in a large group of immunocompetent subjects.

Conclusions

Studies of cancer in people with AIDS have provided unique opportunities to examine the role of the immune system in human cancer. It is clear that immunodeficiency, whether congenital, drug-induced, or related to HIV infection, increases the risk of certain, but not of all, types of cancers.

There is strong and consistent evidence that the immunosuppression associated with HIV infection increases the risk of Kaposi's sarcoma, non-Hodgkin's lymphoma, Hodgkin's disease, squamous cell carcinoma of the conjunctiva, and leiomyosarcoma in children. Most of these cancers are known to be caused by specific herpesviruses. The individuals who develop immunodeficiency-related cancers seem to have characteristics similar to those of immunocompetent individuals who develop the same type of cancer.

Most other cancers show no increase in risk associated with HIV infection, although relatively small increases for rare tumors cannot be excluded. The available evidence suggests that invasive cervical cancer and hepatocellular cancer, both of which are known to be caused by infectious agents, do not appear to be increased markedly in people with AIDS.

Understanding why immunosuppression increases the risk of certain, but not all, cancers that are known to be caused by infectious agents may lead to important insights into the carcinogenic process. In addition, understanding why certain viruses can be found in association with tumors in HIV-seropositive subjects, but not in similar tumors in HIV-seronegative subjects, may aid our understanding of the role of these infections in the etiology of cancer in the general population.

With the prospect of improved survival in HIV-infected people, it is difficult to predict what will happen to cancer incidence in that population. Improved survival should, nevertheless, focus attention on the risk of HIV-associated cancer and may lead to further understanding of the role of infectious agents and the immune system in cancer.

References

(1) Burnet M. Somatic mutation and chronic disease. Br Med J 1965;1:338–42.

(2) Mueller BU, Pizzo PA. Cancer in children with primary or secondary immunodeficiencies. J Pediatr 1995;126:1–10.

(3) Kinlen LJ, Sheil AG, Peto J, Doll R. Collaborative United Kingdom–Australasian study of cancer in patients treated with immunosuppressive drugs. Br Med J 1979;2:1461–6.

(4) Birkeland SA, Storm HH, Lamm LU, Barlow L, Blohme I, Forsberg B, et al. Cancer risk after renal transplantation in the Nordic countries, 1964–1986. Int J Cancer 1995;60:183–9.

(5) Thomas JA, Crawford DH. Epstein-Barr virus associated B-cell lymphomas in AIDS and after organ transplantation. Lancet 1989;1:1075–6.

(6) IARC monograph on the evaluation of carcinogenic risks to humans. Volume 67, Human immunodeficiency viruses and human T-cell lymphotropic viruses. Lyon, France: IARC, 1996.

(7) Barr BB, Benton EC, McLaren K, Bunney MH, Smith IW, Blessing K, et al. Human papilloma virus infection and skin cancer in renal allograft recipients. Lancet 1989;1:124–9.

(8) Starzl TE, Naleshik MA, Porter KA, Ho M, Iwatsuki S, Griffith BP, et al. Reversibility of lymphomas and lymphoproliferative lesions developing under cyclosporin-steroid therapy. Lancet 1984;1:583–7.

(9) Bavinck JN, De Boer A, Vermeer BJ, Hartevelt MM, van der Woude FJ, Claas FH, et al. Sunlight, keratotic skin lesions and skin cancer in renal transplant recipients. Br J Dermatol 1993;129:242–9.

(10) Newton R. A review of the aetiology of squamous cell carcinoma of the conjunctiva. Br J Cancer 1996;74:1511–3.

(11) Grulich A, Wan X, Law M, Coates M, Kaldor J. Rates of non-AIDS defining cancers in people with AIDS. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A18.

(12) Rabkin C. Epidemiology of malignancies other than Kaposi's sarcoma and non-Hodgkin's lymphoma in HIV infection. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A12.

(13) Pollock BH, Jenson HB, McClain KL, Leach CT, Joshi VV, Murphy SB. Risk factors for HIV-related malignancies in children. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A17.

(14) Rabkin CS, Biggar RJ, Horm JW. Increasing incidence of cancers associated with the human immunodeficiency virus epidemic. Int J Cancer 1991; 47:692–6.

(15) Sitas F, Bezwoda WR, Levin V, Ruff P, Kew MC, Hale MJ, et al. Association between human immunodeficiency virus type 1 infection and cancer in the black population of Johannesburg and Soweto, South Africa. Br J Cancer 1997;75:1704–7.

(16) Newton R, Grulich A, Beral V, Sindikubwabo B, Ngilimana PJ, Nganyira A, et al. Cancer and HIV infection in Rwanda [letter]. Lancet 1995;345: 1378–9.

(17) Cote TR, O'Brien TR, Ward JW, Wilson SE, Blattner WA. AIDS and cancer registry linkage: measurement and enhancement of registry completeness. The National AIDS/Cancer Match Study Group. Prev Med 1995; 24:375–7.

(18) Rettig MB, Ma HJ, Vescio RA, Pold M, Schiller G, Belson D, et al. Kaposi's sarcoma-associated herpesvirus infection of bone marrow dendritic cells from multiple myeloma patients. Science 1997;276:1851–4.

(19) Wabinga HR, Parkin DM, Wabwire-Mangen F, Mugerwa JW. Cancer in Kampala, Uganda, in 1989–91: changes in incidence in the era of AIDS. Int J Cancer 1993;54:26–36.

(20) Bassett MT, Chokunonga E, Mauchaza B, Levy L, Ferlay J, Parkin DM. Cancer in the African population of Harare, Zimbabwe, 1990–1992. Int J Cancer 1995;63:29–36.

(21) Newton R, Ziegler JL, Beral V, on behalf of the Uganda Kaposi's sarcoma

study group. A review of the aetiology of squamous cell carcinoma of the conjunctiva: with preliminary results of a case–control study from Kampala, Uganda. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14;A17.
(22) Rabkin CS, Biggar RJ, Baptiste MS, Abe T, Kohler BA, Nasca PC. Cancer incidence trends in women at high risk of human immunodeficiency virus (HIV) infection. Int J Cancer 1993;55:208–12.
(23) Darby SC, Ewart DW, Giangrande PL, Dolin PJ, Spooner RJ, Rizza CR. Mortality before and after HIV infection in the complete UK population of haemophiliacs. U.K. Haemophilia Centre Directors' Organisation. Nature 1995;377:79–82.
(24) Dorffel J. Histogenesis of multiple idiopathic hemorrhagic sarcoma of Kaposi. Arch Dermatol Syph 1932;26:608–34.
(25) Tappero JW, Conant MA, Wolfe SF, Berger TG. Kaposi's sarcoma. Epidemiology, pathogenesis, histology, clinical spectrum, staging criteria and therapy. J Am Acad Dermatol 1993;28:371–95.
(26) Casabona J, Salas T, Salinas R. Trends and survival in AIDS-associated malignancies. Eur J Cancer 1993;29A:877–81.
(27) Ziegler JL, Katongole-Mbidde E. Kaposi's sarcoma in childhood: an analysis of 100 cases from Uganda and relationship to HIV infection. Int J Cancer 1996;65:200–3.
(28) Mbulaiteye SM, Ziegler JL, Katongole-Mbidde E, Tindyebwa D, Marrum L, Newton R, et al. Kaposi's sarcoma study group. Risk factors for childhood Kaposi's sarcoma in Uganda: a case–control study. Abstract 7. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A18.
(29) Whitby D, Howard MR, Tenant-Flowers M, Brink NS, Copas A, Boshoff C, et al. Detection of Kaposi sarcoma associated herpesvirus in peripheral blood of HIV-infected individuals and progression to Kaposi's sarcoma. Lancet 1995;346:799–802.
(30) Rickinson A. The role of herpesviruses in immune deficiency-associated lymphomas. Abstract S17. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A15.
(31) Gao SJ, Kingsley L, Li M, Zheng W, Parravicini C, Ziegler J, et al. KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. Nat Med 1996;2:925–8.
(32) Beral V. Epidemiology of Kaposi's sarcoma. Cancer Surv 1991;10:5–22.
(33) Melbye M, Cook PM, Hjalgrim H, Begtrup K, Simpson GR, Biggar RJ, et al. Risk factors for HHV-8 seropositivity and progression to Kaposi's sarcoma in a cohort of homosexual men, 1981–96. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A16.
(34) Grulich A, Olsen S, Hendry O, Luo K, Cooper D, Zheng W, et al. Route of transmission of Kaposi's sarcoma associated herpesvirus. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A16.
(35) Geddes M, Franceschi S, Barchielli A, Falcini F, Carli S, Cocconi G, et al. Kaposi's sarcoma in Italy before and after the AIDS epidemic. Br J Cancer 1994;69:333–6.
(36) Templeton AC, Buxton E, Bianchi A. Cancer in Kyadondo County, Uganda, 1968–70. J Natl Cancer Inst 1972;48:865–74.
(37) Simpson GR, Schulz TF, Whitby D, Cook PM, Boshoff C, Rainbow L, et al. Prevalence of Kaposi's sarcoma associated herpesvirus infection measured by antibodies to recombinant capsid protein and latent immunofluorescence antigen. Lancet 1996;348:1133–8.
(38) Beral V, Peterman T, Berkelman R, Jaffe H. AIDS-associated non-Hodgkin lymphoma. Lancet 1991;337:805–9.
(39) Moran CA, Tuur S, Angritt P, Reid AH, O'Leary TJ. Epstein-Barr virus in Hodgkin's disease from patients with human immunodeficiency virus infection. Mod Pathol 1992;5:85–8.
(40) Levine PH, Pallesen G, Ebbesen P, Harris N, Evans AS, Mueller N. Evaluation of Epstein-Barr virus antibody patterns and detection of viral markers in the biopsies of patients with Hodgkin's disease. Int J Cancer 1994;59:48–50.

Note

This research was supported by the Imperial Cancer Research Fund, U.K.

Current Perspectives on the Molecular Pathogenesis of Virus-Induced Cancers in Human Immunodeficiency Virus Infection and Acquired Immunodeficiency Syndrome

*Elliott Kieff**

A distinct group of cancers particularly threaten human immunodeficiency virus (HIV)-infected people. Most HIV/acquired immunodeficiency syndrome (AIDS)-associated cancers have a substantial component of viral etiology. Epstein-Barr virus (EBV), Kaposi's sarcoma-associated herpesvirus (HHV8), human papillomavirus (HPV), and HIV have been implicated in the etiology of cancers in AIDS. The molecular mechanisms by which HPV, EBV, HHV8, and HIV persist and cause cancer are summarized. The viral etiology of AIDS-associated cancers is important because pharmacologic and immunologic strategies to prevent or attack persistent or latent virus infection and cell growth transformation may be useful in preventing and treating these cancers. Effective immune attack on latent and persistent virus infection will require enhanced cellular immune responses. Such responses may be achievable through active immunization or by *in vitro* expansion of viral and host specific cytotoxic and helper T lymphocytes. Enhanced knowledge of clinically applied T-cell immunology may also be useful in preventing and treating HIV infection and other opportunistic infections in HIV-infected people. [Monogr Natl Cancer Inst 1998;23:7–14]

Viruses are obligate intracellular parasites that are now recognized to be an etiologic factor in cancers in all human populations [reviewed in *(1,2)*]. These simple organisms have RNA or DNA genomes and have evolved to propagate themselves in human populations. Their causation of cancer is an unusual outcome of an infection that has gone awry in an individual host. Nevertheless, estimates of the frequency of virus-induced cancer worldwide are in the order of 15%–20% of all cancers. In human immunodeficiency virus (HIV) infection or acquired immunodeficiency syndrome (AIDS) (HIV/AIDS), viruses are estimated to cause a significantly higher fraction of cancer. The importance of viruses in HIV/AIDS-associated cancer is due to at least five factors. The first factor is the relatively young age of HIV/AIDS-infected people and the relatively low incidence of non-viral-associated cancers in young people. Second, the acquisition of HIV infection is associated with circumstances that place the individual at increased risk for acquiring other persistent virus infections. Third, some viruses that can cause persistent infections have evolved strategies for evading immune responses, for persisting in a latent state within cells, or for altering cell growth. These strategies can enable virus-infected cells to become malignant. Fourth, HIV infection alters cellular, cytokine, and humoral components of the immune response, decreases antiviral cellular immune responses, increases the burden of persistent virus infections, and thereby predisposes HIV-infected people to develop viral-associated cancers. In being at increased risk for viral-associated cancers because of decreased antiviral immune responses, HIV/AIDS-infected people are similar to other people with decreased cellular immune responses such as organ transplant recipients. Fifth, at the cellular level, HIV infection leads to provirus integration. Integrated proviral DNA can cause disruption or dysregulation of cellular gene expression and thereby initiate or promote oncogenesis. This brief overview will focus on the current status of knowledge of the molecular mechanisms by which viruses persist and effect changes in cell growth that result in cancer in patients with HIV/AIDS. Knowledge of these molecular mechanisms is highly relevant to strategies for the prevention and treatment of HIV/AIDS-associated cancers.

The viruses associated with cancer in HIV/AIDS-infected people are a subset of the viruses associated with cancer in non-HIV-infected people. In non-HIV-infected populations, human papillomaviruses (HPVs) are the initiating etiologic agents in most cervical and anogenital carcinomas, in many laryngeal carcinomas, and in a small fraction of cutaneous carcinomas. Hepatitis B and C viruses (HBV and HCV) are major factors in the etiology of hepatocellular carcinomas (HCC), a prevalent cancer worldwide. Epstein-Barr virus (EBV) is etiologically implicated in nasopharyngeal carcinoma (NPC), a cancer affecting specific populations, in B-cell proliferative disease in immune compromised patients, in lymphomas, in Hodgkin's disease (HD), and in a small fraction of gastric carcinomas. Human T-cell leukemia virus (HTLV-1) is implicated in an uncommon tumor, adult T-cell leukemia (ATL). Human herpesvirus 8 (HHV8, also known as KSHV) is implicated in Kaposi's sarcoma (KS), an unusual tumor in most human populations, but more common

**Affiliation of author:* Programs in Virology, Immunology, and Basic Biological Sciences, Departments of Microbiology, Molecular Genetics, and Medicine, Harvard University, Boston, MA.

Correspondence to: Elliott Kieff, M.D., Ph.D., Programs in Virology, Immunology, and Basic Biological Sciences, Departments of Microbiology, Molecular Genetics, and Medicine, Harvard University, Channing Laboratory, BWH, 181 Longwood Ave., Boston, MA 02115. E-mail: Ekieff@rics.bwh.harvard.edu

See "Note" following "References."

© Oxford University Press

among some Mediterranean and African populations, in multicentric Castleman's disease, and in body cavity B-cell lymphomas (BCBLs). HPV, HBV, HCV, HHV8, and perhaps, to a small extent, HTLV-1 are more common in people exposed to blood, blood products, or sexually transmitted diseases and are therefore more prevalent in HIV/AIDS, while EBV is highly prevalent in all populations. In HIV/AIDS, HHV8 is associated with KS and BCBL, and EBV is associated with central nervous system and peripheral lymphomas, HD, and leiomyomas, all of which occur with increased frequency. In contrast, despite attributable viral etiology, HCC, cervical cancer, and ATL appear to be no more common than in the general population. HPV infection appears to be intermediate in increased association with cancer in HIV/AIDS in that while cervical and anogenital cancers are not much increased in incidence, HPV-associated premalignant lesions of all types are more common in patients with HIV/AIDS. The failure to observe an increased frequency of cervical and anogenital invasive cancers in AIDS may be due to the previously short life span of HIV-infected people and the usual long interval between HPV infection and cancer. Similarly, the long interval between HBV or HCV infection and HCC or between HTLV-1 and ATL is likely to be a factor in the low incidence of HCC and ATL in HIV/AIDS. However, the low incidence of HCC and ATL in HIV/AIDS may also be due to specific effects of HIV infection. HIV infection depletes CD4+ T lymphocytes, which are critical for enhanced CD8+ T-cell immunc responses, and changes in cytokine levels shift the immune responses away from natural killer (NK) and CD8+ lymphocytes and toward B lymphocytes. Since CD8+ lymphocyte-mediated liver injury and regeneration are important in HBV or HCV induction of HCC *(3)* and interleukin 2 secretion from CD4+ cells is critical for the proliferation of HTLV-1-infected cells *(4)*, HIV infection may actually decrease the frequency of cancer associated with HBV, HCC, or HTLV-1 infection. Because HCC and ATL are not commonly associated with HIV/AIDS, they will not be further discussed in this review. Also, viral-associated benign proliferations that are important in HIV/AIDS, such as Mulluscum Contagiosum (due to the mulluscum poxvirus) and oral hairy leukoplakia (associated with EBV), are only mentioned at this point.

Human Papillomavirus

HPV 16, HPV 18, and other high-risk papillomaviruses are strongly associated with cervical, penile, and anal cancers and HPV 5 and HPV 8 have been associated with widespread epidermal, dysplastic lesions, and some carcinomas in HIV-infected and noninfected people *(5–9)*. Important aspects of the molecular mechanisms through which HPVs cause persistent infection and cancer are now partially understood. At the cellular level, HPVs infect basal epithelial cells and persist as an episome in the latently infected cell (Fig. 1). It is uncertain whether any virus-encoded proteins are expressed in infected basal epithelial cells. Basal epithelial cells replicate and some progeny differentiate into nondividing parabasal cells that further differentiate into keratinocytes. Viral replication ensues in infected differentiating keratinocytes. The HPV early promoter governs transcription of the seven HPV early genes (E1–E7). The high-risk HPV E7 proteins bind to the cellular retinoblastoma protein and thereby release the cellular transcription factor E2F. E2F is then free to

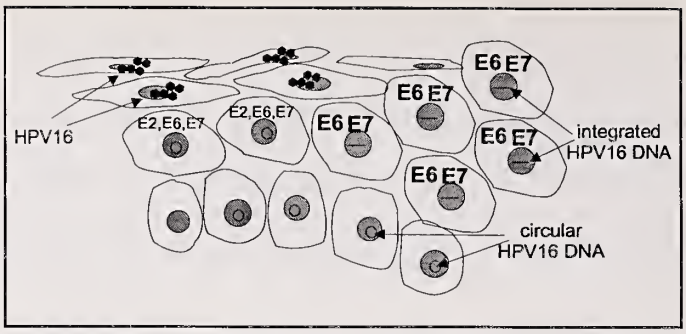


Fig. 1. Schematic depiction of HPV 16-persistent infection and replication in epithelial cells on the left with a focus of potentially oncogenic integration and high level E6 and E7 expression on the right.

up-regulate the expression of cellular genes that are important for viral and cellular DNA synthesis. High-risk HPV E6 proteins cause the degradation of p53. p53 would otherwise be induced by HPV infection, inhibit cellular cyclin dependent kinases, and cause apoptosis. The HPV E1 and E2 proteins enable viral DNA replication. When the E2 protein accumulates to high levels, the viral early promoter is down-regulated *(10)* and an as yet unidentified promoter that transcribes messenger RNAs (mRNAs) for the late viral structural proteins is presumed to be turned on.

HPV DNA integration into chromosomal DNA is not part of the normal strategy by which HPV persists in cells. Integration is the first major step away from the events of normal persistent HPV infection and toward cancer. In carcinoma tissue or carcinoma cell lines in culture, HPV DNA is usually integrated with the E2 open-reading frame interrupted so that the E2 protein is not expressed or it is unable to down-modulate the early promoter *(10–13)*. Integrations that release the HPV early promoter from the repressive effects of E2 result in continuous high-level expression of E6 and E7. While normal levels of E6 and E7 expression in an infected, differentiating keratinocyte may enable HPV to synthesize its own DNA, high-level E6 and E7 expression from a high-risk HPV in a cell that is not fully committed to terminal differentiation has immortalizing effects *(14)*. E7 has several functions, including an ability to associate with the part of the retinoblastoma protein that would otherwise bind to and inactivate E2F transcription factors *(15,16)*. The freeing of E2F results in activation of the many E2F-responsive cellular genes whose expression enables a resting cell to enter cycle, traverse G_1, and enter S phase *(17)*. In contrast to E7 that acts in the nucleus, HPV E6 acts in the cytoplasm where it binds to several cellular proteins. One important interaction is with a ubiquitin ligase *(18)*. The E6-associated ubiquitin ligase catalyzes the ubiquitination of the p53 tumor suppressor gene and its subsequent destruction by cellular proteosomes. By inducing the degradation of p53, E6 prevents p53's tumor-suppressive effects, including the induced expression of cyclin-dependent kinase (CDK) inhibitors and the induction of apoptosis. High-risk HPVs characteristically have E6 and E7 proteins that are significantly more active in targeting p53 and the retinoblastoma protein than E6 and E7 of low-risk HPVs. Although HPV DNA appears to randomly integrate into cellular chromosomes, some integration events may promote tumorigenicity by enhancing transcription of neighboring cellular oncogenes. One example is the integration of HPV DNA near c-myc in a cervical cancer cell line. c-myc is also amplified in some tumors *(11)*.

HPV infection, integration of the HPV genome into cellular DNA, and overexpression of high-risk and transforming E6 and E7 proteins are important steps in a multistep carcinogenic process. In most populations, 10%–30% of adults are infected with high-risk HPVs. HPV integration is unusual and an integration that results in HPV early promoter up-regulation is a key oncogenic event, since each tumor is marked by a single distinctive integration. Progression through low- and high-grade intraepithelial neoplasia to invasive carcinoma appears to usually require several decades. Most infected people never even progress to low-grade intraepithelial neoplasia, and invasive carcinoma is a very infrequent outcome. Many invasive carcinomas have a loss of heterozygosity at 3p, implicating this site and the FHIT gene as being important for progression *(19,20)*. Chromosome 11 has also been implicated through its activity in suppressing HPV gene expression and oncogenicity *(5)*.

One reason that most HPV-infected people never progress to invasive cervical cancer is that HPV infection can be cleared by immune mechanisms. Despite the somewhat inefficient overall immune surveillance in epithelial tissue, humoral and cell-mediated immune responses contain and eventually eliminate most HPV infections. Interferon appears to be an important component of the immune response because of its immune regulatory and antiviral activities. Nevertheless, the frequent persistence of even benign HPVs for months or years and the prolonged persistence of high-risk HPVs for decades in some people are compatible with the notion that HPV-encoded proteins may facilitate immune evasion. However, even high-risk HPVs are usually eventually cleared by the normal immune response. The immune-suppressive effects of HIV infection appear to enable longer term and higher level HPV persistence, placing HIV- and HPV-coinfected people at increased theoretic risk for cervical or anogenital cancer *(10)*. If contemporary antiretroviral therapies have their expected effects in delaying HIV disease progression and in forestalling HIV-associated mortality, HPV infections are likely to have a greater cumulative effect on cancer incidence in HIV-infected people because of the longer, at-risk exposure time. Novel strategies to prevent or treat HPV infection may be more important in this era of improved antiretroviral therapies.

Epstein-Barr Virus

EBV infection [reviewed in *(21,22)*] is similar to HPV infection in that the viral genes have a direct role in oncogenesis, but it is strikingly different in three respects. First, while the EBV genes involved in oncogenesis perform the same functions in normal latent virus infection, HPV E6 and E7 are important for early stages of viral replication and are not known to be expressed in latently infected cells. Second, high-risk E6 and E7 genes only become oncogenic as a result of integration and overexpression in cells that are not terminally differentiated. In contrast, the EBV genes important for oncogenesis are normally expressed in latent infection. They enable EBV to expand the number of latently infected cells and establish persistence in the normal host. Third, since the EBV genes associated with oncogenesis are part of the normal life cycle of the virus, their ability to cause cell proliferation must be limited so as to not cause the premature demise of the host and the virus. This later balance is struck by the reliance of the virus on the normal innate and immune response to viral infection. These responses strongly contain virus-infected cell proliferation. Because of its dependence on innate and acquired immunity, EBV infection is rarely oncogenic in people with normal immune function and is a more important cause of cancer in severely immune-compromised people, especially people with advanced HIV infection or AIDS.

EBV infection is usually spread by saliva and virus replicates in the oropharyngeal epithelium. The molecular events of intracellular replication are similar to those of other herpesviruses. In primary infection, EBV spreads from epithelial cells to B lymphocytes. In B lymphocytes, the EBV genome circularizes and expresses two nuclear proteins (EBNA-LP and -2). EBNA-LP and EBNA-2 turn on the expression of some cell genes and up-regulate their own promoter and an upstream promoter, thereby increasing transcription through the EBNA-2 polyadenylation site with downstream transcription of four additional nuclear protein (EBNA-3A, -3B, -3C, and -1) mRNAs. EBNA-LP and EBNA-2 also turn on the transcription of mRNAs encoding two integral membrane proteins (LMP-1 and -2). EBV latent infection-associated gene expression results in rather uniform progression of the infected B lymphocyte from G_0 into G_1 and S phase, the expression of small RNAs (EBERs) from the EBV genome, and continuous cell proliferation. The EBNAs and LMPs are not related to genes of other herpesvirus.

In vitro, EBV infection of B lymphocytes results in the expression of the EBNAs and LMPs and the establishment of long-term lymphoblastoid cell lines (LCLs). Similar EBV-infected LCLs result from the cultivation, *in vitro,* of peripheral blood B lymphocytes from EBV-seropositive people. Injection of EBV into cottontop tamarins causes an acute fatal polyclonal lymphoproliferative disease *(23)*, similar to that seen in severely immune-compromised humans with primary or even latent EBV infection. In more normal people, cells expressing EBNAs and LMPs engender natural killer and EBV-specific, major histocompatibility complex (MHC) class I-restricted, cytotoxic CD8+ T-cell responses *(24–26)*. The EBNAs in particular, except for EBNA-1, have multiple epitopes that are recognized in the context of common class I determinants. The EBNAs and LMP1 also induce the expression of adhesion molecules, rendering the cell susceptible to T-cell adherence and cytocidal effects. As a consequence of immune responses by normal people to primary EBV infection, the number of proliferating virus-infected B lymphocytes in the peripheral blood rapidly declines to a level of one infected B lymphocyte in 10^{-5}or 10^{-6}. However, cytotoxic T lymphocytes specific for epitopes from five of the EBNAs and the two LMPs persist forever, indicating that cells expressing the EBNAs and LMPs are at least intermittently present in the normal host.

Some EBV-infected B lymphocytes switch to a latent infection in which only EBNA-1 is expressed. These cells are apparently not proliferating and EBNA-1 is not usually recognized by immune CD8+ lymphocytes, so that cells expressing only EBNA-1 are immunologically indistinguishable from normal B lymphocytes. EBNA-1 appears to escape immune surveillance because it has a *cis*-acting glycine alanine repeat sequence that inhibits its processing through proteosomes *(27)*. The expression of EBNA-1 in the absence of other EBNAs and LMPs was first described by Rickinson et al. *(22)* in studies of Burkitt's lymphoma cell lines and tissue and has led to the working hypothesis

that there may be immunologic selection against EBV gene expression in the later stages of lymphoma evolution.

Long after primary EBV infection, EBV is closely associated with endemic Burkitt's lymphoma, sporadic Burkitt's lymphoma, T-cell lymphomas, HD, and anaplastic NPC in normal hosts. In HD and NPC, EBNA-1 and LMPs are expressed in the absence of other EBNAs. LMP-1 appears to be a central effector of altered cell growth in lymphocytes or epithelial cells.

Recombinant EBV-based reverse genetic analyses have shown that EBNA-2, EBNA-LP, EBNA-3A, EBNA-3C, and LMP-1 are critical or essential for virus-mediated B-cell growth transformation *(28)*. EBNA-1 is also presumed to be important, since EBNA-1 binds to a site upstream of the EBNA promoter, thereby enhancing transcription and creating a functional origin for replication of the EBV episome in cellular S phase *(29)*. LMP-2, EBNA-3B, small non-polyadenylated RNAs (EBERs), and a long highly spliced mRNA (BARF0), which are also expressed in latent lymphocyte infection, are unimportant for primary B-lymphocyte growth transformation. Their role in latent EBV infection is uncertain except for LMP-2. LMP-2 has a critical role in rendering the transformed cell not susceptible to activation signals that would otherwise result in the activation of lytic EBV infection. Most of the rest of the viral genome, including EBV-encoded IL10 and Bcl2 analogues that are ordinarily expressed in lytic infection, have also formally been shown to be unimportant in the conversion of resting B lymphocytes into LCLs and subsequent growth into tumors in SCID mice.

Recent biochemical and reverse genetic analyses are compatible with the hypothesis that most of EBV's effects on cell growth are mediated by virus-encoded proteins that usurp control of the notch and CD40 signaling pathways. Early biochemical work established EBNA-2 as a transactivator of expression of the B-lymphocyte activation marker, CD23, and of LMP-1 transcription. Subsequent recombinant EBV reverse genetic analyses defined three critical components of the EBNA-2 open-reading frame. The first is simply several codons that encode prolines near the amino terminus of EBNA-2. These codons are the core of a dimerization element. The second component is codons 280–337 that mediate interactions with PU.1, a B lymphocyte and macrophage-specific "ets" family transcription factor, and with RBP-Jk, a transcription factor that has been genetically implicated in notch-mediated neural development in *Drosophila*. EBNA-2 in an EBV-transformed lymphoblast is strongly associated with RBP-Jk and a substantial fraction of the cellular RBP-Jk is associated with EBNA-2 *(30)*. The third essential EBNA-2 component is codons 424–464, which encode an acidic activator. This acidic activator can interact with TFIIB, TAF40, and two subunits of TFIIH and extensively associates with a novel coactivating protein, p100, and with citric acid lyase, a major acetyl-CoA donor. P100 can interact with two subunits of TFIIE. P100 also has a domain that is functionally equivalent to the major carboxyl terminal negative regulatory domain of the c-myb protein *(31)*. Recent experiments indicate that EBNA-LP coactivates transcription along with EBNA-2. EBNA-LP is not an independent transactivator, but the localization of the EBNA-2 acidic domain near a promoter enables EBNA-LP to strongly coactivate transcription from that promoter *(32)*.

Surprisingly, EBNA-3A and EBNA-3C (as well as EBNA-3B, which is not critical for growth transformation) also associate with RBP-Jk and regulate expression of specific virus and cell genes with RBP-Jk sites. The central role of RBP-Jk in EBV regulation of viral and cellular gene expression, evidence that notch activation is a cause of T-cell leukemias *(33)*, and evidence that RBP-Jk mediates notch effects, are all compatible with the hypothesis that RBP-Jk may mediate the regulation of transcription of some cellular gene(s) that are important in lymphocyte growth control.

EBV uses the EBNAs and RBP-Jk to precisely regulate transcription of its oncogene, LMP-1 (Fig. 2). LMP-1 expression in Rat1 cells causes loss of contact inhibition, anchorage-independent growth, and tumorigenicity in nude mice. In lymphocytes, LMP-1 enhances bcl-2 expression, activates NF-kB and c-jun, and induces most of the activation and adhesion molecules that are activated by EBV infection or by antigen and T-cell help. T cells express CD40 ligand, and the T-cell CD40 ligand/B cell CD40 receptor complex is a key component of T-cell help.

LMP-1 has six hydrophobic transmembrane domains that enable it to constitutively aggregate in the plasma membrane and are essential for EBV transformation of B cells into LCLs (Fig. 3). The observation that a mutation in the LMP-1 transmembrane domains that results in diffuse expression in the plasma membrane has a nontransforming phenotype indicates that aggregation is linked to transformation, most likely because aggregation enables the cytoplasmic domains of several LMP1 molecules to locally concentrate next to the plasma membrane as though they were the cytoplasmic domains of a growth factor receptor that had encountered ligand. The LMP1 amino terminal cytoplasmic domain is not critical, but the carboxyl terminal cytoplasmic domain has two important components: a proximal component that is sufficient for initial immortalization (transformation effector site 1, TES1) and a distal domain that enables efficient LCL outgrowth (TES2) *(28,34,35)*. TES1 and TES2 mediate 30% and 70% of the NF-kB activation from LMP-1, respectively. Since TES1 only mediates a small component of NF-kB activation, but is both necessary and sufficient for initial lymphocyte immortalization, NF-kB can only be a component of TES1 LMP1-mediated growth transformation. Two proteins that interact with LMP1 were identified in a yeast-two hybrid screen and in a screen for proteins induced by EBV infection. These proteins are highly homologous to proteins identified as mouse tumor necrosis factor receptor II (TNFRII)-associated factors or

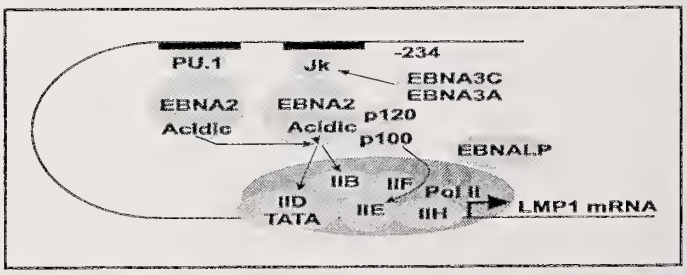


Fig. 2. Schematic diagram of the regulation of the EBV LMP1 promoter by EBV nuclear proteins in latently infected B lymphocytes. Not shown is EBNA-1 that also contributes to LMP1 up-regulation through its binding to the ori-p element, which is separated by about 10 kilobase pairs from the LMP1 promoter.

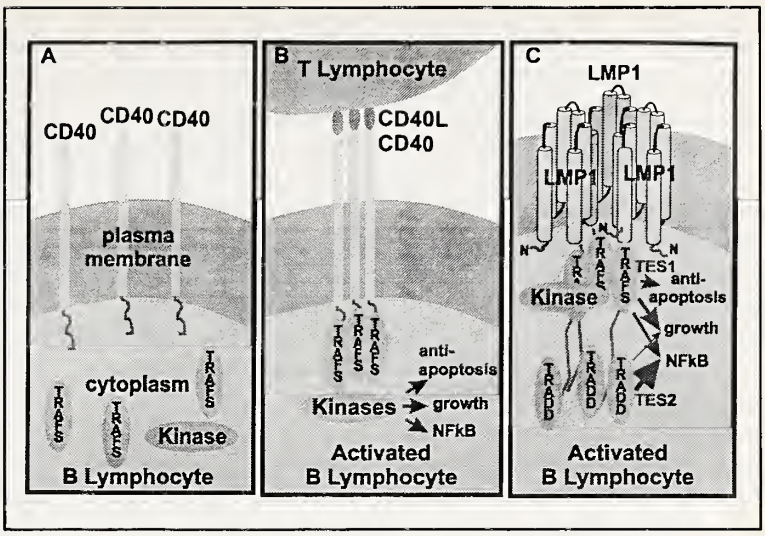


Fig. 3. Schematic diagram of signal transduction from EBV LMP1 and its similarity to ligand-activated CD40 and other activated tumor necrosis factor receptors. LMP1 is constitutively activated because of the six hydrophobic transmembrane domains that cause LMP1 to noncovalently aggregate in the plasma membrane.

TRAFs. LMP1 is similar to activated CD40 in its effects on B-lymphocyte growth and CD40 is also a TNFR that associates with TRAFs, implicating TRAFs in signaling from LMP1 (Fig. 3). In EBV-transformed B lymphoctes, LMP1 is constitutively strongly associated with TRAF1 and TRAF3 and is also associated to a lesser extent with TRAF2 *(36)*. Such cells have a high level of TRAF1/2 heterodimers that mediate NF-kB activation from TES1. The pathway appears to be similar to TNFRII, LT-beta receptor, or CD40 ligand-dependent activation of NF-kB through TRAF2. Furthermore, like CD40, LMP1 activates SAPK and c-jun through TRAF2. Thus, the LMP-1 transmembrane domains mediate constitutive aggregation in the plasma membrane, enabling the LMP1 cytoplasmic domain to be associated with TRAFs and mediate NF-kB and c-jun activation, as well as other less well-characterized effects important for cell growth stimulation. Deletion of DNA encoding the TRAF binding site from the LMP1 gene in recombinant EBVs results in a null phenotype for resting B-lymphocyte growth transformation.

Recent experiments *(35)* indicate that TES2 maps to the last three residues of LMP1. High-level NF-kB activation is also mediated by this site, providing genetic evidence that high-level NF-kB activation is important in LMP1's effects. The TNF receptor death domain protein (TRADD) uniquely interacts with TES2 and is constitutively associated with LMP1 in EBV-transformed B lymphocytes. These data implicate TRADD in LMP1-mediated growth transformation and NF-kB activation, underscoring the degree to which LMP1 mimics a constitutively activated TNFR. Curiously, LMP1 engagement of TRADD does not appear to effect cell death, whereas bcl-2 and NF-kB activation by LMP1 protect EBV-infected cells from cell death.

Most of the central nervous system lymphomas that occur late in AIDS and some of the peripheral lymphomas are characterized by the expression of the full range of EBNAs and LMPs. The occurrence of EBV-associated lymphomas is consistent with earlier studies, indicating that the number of lymphocytes latently infected with EBV increases during HIV infection. Waning cytotoxic T-cell immunity is likely to be a critical factor in the increased abundance of EBV-infected cells in the peripheral blood and the subsequent emergence of lymphoma. Changes in cytokine levels and the shift in cytokine balance toward B lymphocyte up-regulation may also have a role.

The fact that many of the EBV-associated lymphomas that occur late in AIDS are similar in viral and cellular gene expression to resting B lymphocytes transformed by EBV infection *in vitro* and to B lymphocytes in the EBV-associated oligoclonal lymphoproliferative disease that can occur with primary EBV infection in organ transplant recipients who are receiving high dose immune-suppressive therapy has two important implications. First, the lymphoma cells may still be critically dependent on EBV gene expression for their proliferation. Second, the ability of cytotoxic T lymphocytes from normal EBV-infected people to kill latently infected cells through EBNA- or LMP-derived epitopes in the context of common class I MHC molecules can be exploited for the prevention or treatment of EBV-induced lymphoproliferative diseases in which these viral proteins are expressed.

Clinical research with human T-cell immunotherapy has already achieved some promising results in post-transplant patients *(37,38)*. For bone marrow transplant recipients, infusion of donor peripheral blood mononuclear cells or of donor T-cell lines derived through exposure of T lymphocytes to autologous EBV-transformed B lymphocytes has resulted in the regression or the prevention of EBV-induced lymphoproliferative disease. In initial studies, donor T cells appear able to survive, hone to sites of proliferating EBV-infected cells, effect killing, and proliferate to a limited extent. Similar approaches might be beneficial in patients with AIDS. Since the majority of patients with AIDS escape EBV-associated lymphomas, efforts to increase cytotoxic T-cell responses by active immunization or by expansion of EBV-reactive autologous T lymphocytes, *in vitro,* and reinfusion might confer protection against EBV-associated lymphoproliferations.

Many EBV-associated cancers in patients with HIV/AIDS are likely to be EBV-initiated and to have subsequent chromosomal changes that advance the malignant phenotype. Subsequent chromosomal changes are well described in Burkitt's lymphomas and also characterize much of the spectrum of lymphoproliferative disease that occurs in post-transplant lymphomas and HIV/AIDS *(39)*. The most frequent change is a c-myc translocation leading to dysregulated c-myc expression. Activation of c-myc does not necessarily indicate independence of EBV gene expression for continued cell growth, and EBV gene expression

in malignant cells provides a basis for immune attack on the tumor cell.

Other EBV-associated cancers that occur with a higher frequency in HIV/AIDS include HD and leiomyosarcomas. EBV-associated HD is characterized by EBNA-1, LMP2, and high-level LMP1 expression and by the absence of expression of other EBNAs. With regard to HD, the similarities of LMP1's biochemical effects to those of activated CD30, the high level of CD30 expression on EBV-associated as well as EBV non-associated HD tumor cells, and the presence of T cells that are likely to express CD30 ligand in HD tissues are compatible with the hypothesis that CD30 and LMP1 are parallel signaling pathways important in HD tumor cell growth. EBV-infected HD cells may also be subject to immune attack *(40)*. While lymphomas and leiomyomas are substantially increased in patients with HIV/AIDS and HD is marginally increased, NPC, the most common EBV-associated cancer worldwide, is uncommon in patients with AIDS. As with cervical cancer and HCC, the failure to observe an increased incidence of NPC in HIV/AIDS may be related to the long interval between EBV infection and NPC.

HHV8 (KSHV)

The discovery of HHV8 (or KSHV), a second human gamma herpesvirus, in human KS tissue by Chang et al. *(41)* is a breakthrough in understanding the etiology of KS. Previously, cytokines and HIV TAT had been implicated in the genesis of KS *(42)*. However, TAT levels have been difficult to measure in tissues and the epidemiology of KS in HIV-infected people is most compatible with a second sexually transmitted agent *(43)*. Importantly, KS tumors from patients with HIV/AIDS or even from patients with familial KS almost always contain HHV8 DNA *(44)*. Furthermore, KS spindle cells that express CD54, an endothelial cell marker, also have RNA from a restricted portion of the HHV8 genome, indicating that these cells are latently infected with HHV8 *(45)*. A few spindle cells also contain RNA encoding a late viral structural protein, which is compatible with similar findings in EBV-associated lymphoproliferative lesions. Moreover, in patients with HIV/AIDS, antibody to KSHV is a predictor of subsequent KS. Although there has been considerable controversy about the frequency of HHV8 DNA in semen and peripheral blood and about the frequency of HHV8 antibody in various populations, more recent data are consistent with the notion that HHV8 is a sexually transmitted virus that may infect 10% of the general population and a higher fraction of the HIV/AIDS population *(46,47)*. Thus, there is now considerable evidence implicating HHV8 in KS. However, relatively little is known about the mechanisms by which HHV8 might effect cell growth. Additional evidence that would strengthen the case for HHV8 in the etiology of KS and BCBLs *(48)* could come from studies demonstrating that HHV8 can transform cells in culture or induce tumors in heterologous species.

The etiology of BCBL may be complex. An HHV8 ORF73-encoded viral nuclear protein is expressed in BCBL cells *(49)*. The ORF73 latency-associated HHV8 protein has been termed LANA for latency-associated nuclear antigen. However, EBV is also usually present in BCBLs and may have a role in oncogenicity.

BCBL cells can be grown in culture and EBV-free BCBLs are at the present time the best source of HHV8 DNA or proteins. HHV8 replication can be induced in BCBL cells and large quantities of HHV8 are produced from some cell lines *(50)*. So far, only limited replication has been described in primary B lymphocytes and only abortive infection has been described in 293 cells.

Sequence analysis of the HHV8 genome has revealed homology to other herpesviruses, particularly to the rhadinovirus subgroup of the gamma herpesviruses. While EBV has been the most intensely studied gamma herpesvirus, much is also known about the somewhat more closely related rhadinovirus, herpesvirus *saimiri* (HVS). HVS is a T-lymphotropic gamma herpesvirus that can induce T-cell lymphomas in some heterologous species, but has no such effect on its natural host *(51,52)*. HVS has been investigated using biochemical and molecular genetic approaches and may offer some instructive precedent for how HHV8 may cause persistent infection and effect cell growth (Fig. 4). HVS has two putative transforming genes near the left end of its genome. The H. *saimiri* transforming protein and tyrosine kinase interacting protein encoding genes, STP and TIP, vary among HVS strains. Recombinant HVS reverse genetic analyses indicate that STP and TIP are essential for HVS lymphomagenesis. STP up-regulates the Ras pathway *(53)*, whereas TIP down-regulates Lck and activates NF-kB *(54)*. TIP down-regulation of Lck appears to be unimportant for transformation, while TIP-mediated NF-kB activation is likely to be important in lymphomagenesis. Whether the HVS-encoded cyclin homolog, a superantigen encoding open-reading frame, a bcl-2 homolog, and an IL-17 homolog have any role in HVS-induced lymphomagenesis is uncertain. HVS also has several open-reading frames that are likely to be important in escape from immune surveillance and similar open-reading frames are present in HHV8.

HHV8 has 75 open-reading frames that are homologous to HVS, which have been designated ORF1-ORF75, and at least 15 novel open-reading frames that have been designated K1–15 *(51)*. K1 is positioned at the site of HVS STP and TIP and is also variable in sequence among HHV8 strains. Two other novel HHV8 genes could also be important in transformation. HHV8 K2 encodes a functional IL-6 homolog that could be important in endothelial cell growth and HHV8 K13 encodes an inhibitor of FADD-activated apoptosis *(52,55–58)*. Other HHV8 open-reading frames that have homologs in HVS of unproven significance for HVS-induced T-cell transformation are HHV8 ORF72, the cyclin homolog, and ORF16, a bcl-2 homolog. However, the only HHV8 protein known to be expressed in latently infected cells is the ORF73-encoded protein, LANA. A

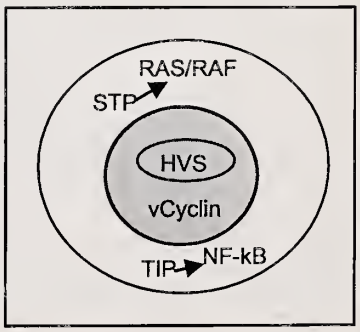


Fig. 4. Schematic diagram of the molecular mechanisms by which herpesvirus *saimiri* (HVS) induces T-cell lymphomas in heterologous primate species. Genetic analyses have confirmed the importance of STP and TIP in T-cell lymphomagenesis. The role of the HVS cyclin D homolog in T-cell lymphomagenesis has not been evaluated, because the gene is essential for virus replication.

lytic infection-associated protein such as the HHV8 IL-6 homolog could be important for KS endothelial cell growth if it is expressed in large amounts from lytically infected cells and such cells are sufficiently prevalent to sustain high cytokine levels. Persistent lytic poxvirus infections, for example, cause hypertrophic cutaneous lesions through secretion of epidermal cell growth factor-related and immune modulating factors. Poxviruses that cause hypertrophic lesions appear to be unique in that infection is highly localized, all infected cells are persistently lytically infected, and there is almost a complete blockade of an effective immune response.

Human Immunodeficiency Virus

The frequency with which HIV itself causes cancer is uncertain and only limited data are available. The positive data are from analyses of non-B-cell lymphomas of all types that were HIV p24 positive *(59)*. Of 22 specimens analyzed in two studies, 18 were positive for HIV proviral DNA by inverse polymerase chain reaction. Of the 18 cases in which proviral DNA was detected, 10 were integrations of HIV upstream of the c-fes gene. Proviral integration upstream of c-fes was found in one T-cell lymphoma. However, almost all of the other integrations were in macrophage-like cells and not in the tumor cells. These findings have led to the hypothesis that HIV integration near c-fes might result in up-regulation of a cytokine that might contribute to lymphomagenesis. Given the frequent finding of defective proviral integrations in HIV infection overall and the substantial activity of the HIV promoter, particularly in cells with high basal NF-kB activation, HIV may be a more significant factor in HIV/AIDS-associated cancer than is currently appreciated.

References

(1) Poeschla E, Wong-Staal F. Etiology of cancer: RNA viruses. In: DeVita VT Jr, Hellman S, Rosenberg SA, editors. Cancer: principles and practice of oncology. Philadelphia: Lippincott-Raven, 1997:153–67.

(2) Howley PM, Ganem D, Kieff ED. Etiology of cancer: DNA viruses. In: DeVita VT Jr, Hellman S, Rosenberg SA, editors. Cancer: principles and practice of oncology. Philadelphia: Lippincott-Raven, 1997:168–84.

(3) Chisari FV, Ferrari C. Hepatitis B virus immunopathogenesis. Ann Rev Immunol 1995;13:29–60.

(4) Gootenberg JE, Ruscetti FW, Mier JW, Gazdar A, Gallo RC. Human cutaneous T cell lymphoma and leukemia cell lines produce and respond to T cell growth factor. J Exp Med 1981;154:1403–18.

(5) zur Hausen H. Papillomavirus infections: a major cause of human cancer. BBA Reviews on Cancer 1966;1288(2):F55–78.

(6) Ostrow RS, Bender M, Niimura M, Seki T, Kawashima M, Pass F, et al. Human papillomavirus DNA in cutaneous primary and metastasized squamous cell carcinomas from patients with epidermodysplasia verruciformis. Proc Natl Acad Sci U S A 1982;79:1634–8.

(7) Durst M, Gissmann L, Ikenburg H, zur Hausen H. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proc Natl Acad Sci U S A 1983;80:3812–5.

(8) Iftner T, Bierfelder S, Csapo Z, Pfister H. Involvement of human papillomavirus type 8 genes E6 and E7 in transformation and replication. J Virol 1988;62:3655–61.

(9) Palefsky JM, Barrasso R. HPV infection and disease in men. Obstet Gynecol Clin North Am 1996;23:895–916.

(10) Thierry F, Yaniv M. The BPV1-E2 trans-acting protein can be either an activator or a repressor of the HPV18 regulatory region. EMBO J 1987;6: 3391–7.

(11) Durst M, Croce CM, Gissmann L, Schwarz E, Huebner K. Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. Proc Natl Acad Sci U S A 1987;84:1070–4.

(12) Sehwarz E, Freese UK, Gissmann L, Mayer W, Roggenbuck B, Stremlau A, et al. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. Nature 1985;314:111–4.

(13) Romanczuk H, Howley PM. Disruption of either the E1 or E2 regulatory region of human papillomavirus type 16 increases viral immortalization capacity. Proc Natl Acad Sci U S A 1992;89:3159–63.

(14) Munger K, Phelps W, Bubb V, Howley PM, Schlegel R. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. J Virol 1989; 63:4417–21.

(15) Munger K, Werness BA, Dyson N, Phelps WC, Harlow E, Howley PM. Complex formation of the human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. EMBO J 1989;8:4099–105.

(16) Chellappan S, Kraus VB, Kroger B, Munger K, Howley PM, Phelps WC, et al. Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between the transcription factor E2F and the retinoblastoma gene product. Proc Natl Acad Sci U S A 1992;89:4549–53.

(17) Johnson DG, Schwarz JK, Cress WD, Nevins JR. Expression of the transcription factor E2F1 induces quiescent cells to enter S phase. Nature 1993;365:349–52.

(18) Scheffner M, Huibregtse JM, Vierstra RD, Howley PM. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell 1993;75:495–505.

(19) Yokota J, Tsukada Y, Nakajima T, Gotoh M, Shimosato Y, Mori N, et al. Loss of heterozygosity on the short arm of chromosome 3 in carcinoma of the uterine cervix. Cancer Res 1989;49:3598–601.

(20) Ohta M, Inoue H, Cotticelli MG, Kastury K, Baffa R, Palazzo J, et al. The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. Cell 1996;84:587–97.

(21) Kieff ED. Epstein-Barr virus and its replication. In: Fields B, Knipe D, Howley PM, editors. Fields virology. Philadelphia: Lippincott-Raven, 1996:2343–96.

(22) Rickinson A, Kieff ED. Epstein-Barr virus. In: Fields B, Knipe D, Howley PM, editors. Fields virology. Philadelphia: Lippincott-Raven, 1996:2397.

(23) Shope T, Dechairo D, Miller G. Malignant lymphoma in cottontop marmosets after inoculation with Epstein-Barr virus. Proc Natl Acad Sci U S A 1973;70:2487–91.

(24) Tomkinson BE, Maziarz R, Sullivan JL. Characterization of the T cell-mediated cellular cytotoxicity during acute infectious mononucleosis. J Immunol 1989;143:660–70.

(25) Khanna R, Burrows SR, Kurilla MG, Jacob CA, Misko IS, Sculley TB, et al. Localization of Epstein-Barr virus cytotoxic T cell epitopes using recombinant vaccinia: implications for vaccine development. J Exp Med 1992;176:169–76.

(26) Murray RJ, Kurilla MG, Brooks JM, Thomas WA, Rowe M, Kieff ED, et al. Identification of target antigens for the human cytotoxic T cell response to Epstein-Barr virus (EBV): implication for the immune control of EBV-positive malignancies. J Exp Med 1992;176:157–68.

(27) Levitskaya J, Coram M, Levitsky V, Imreh S, Steigerwald-Mullen P, Klein G, et al. Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. Nature 1995;375:685–8.

(28) Izumi KM, Kaye KM, Kieff ED. The Epstein-Barr virus LMP1 amino acid sequence that engages tumor necrosis factor receptor associated factors is critical for primary B lymphocyte growth transformation. Proc Natl Acad Sci U S A 1997;94:1447–52.

(29) Yates J, Warren N, Reisman D, Sugden B. A *cis*-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. Proc Natl Acad Sci U S A 1984;81: 3806–10.

(30) Johannsen E, Miller CL, Grossman SR, Kieff ED. EBNA-2 and EBNA-3C extensively and mutually exclusively associate with RBPJkappa in Epstein-Barr virus-transformed B lymphocytes. J Virol 1996;70:4179–83.

(31) Dash AB, Orrico FC, Ness SA. The EVES motif mediates both intermolecular and intramolecular regulation of c-Myb. Genes Dev 1996;10: 1858–69.

(32) Harada S, Kieff ED. Epstein-Barr virus nuclear protein LP stimulates EBNA-2 acidic domain-mediated transcriptional activation. J Virol 1997; 71:6611–8.

(33) Pear WS, Aster JC, Scott ML, Hasserjian RP, Soffer B, Sklar J, et al. Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. J Exp Med 1996;183:2283–91.

(34) Mosialos G, Birkenbach M, Yalamanchili R, VanArsdale T, Ware C, Kieff ED. The Epstein-Barr virus transforming protein LMP1 engages signaling proteins for the tumor necrosis factor receptor family. Cell 1995;80: 389–99.

(35) Izumi KM, Kieff ED. The Epstein-Barr virus oncogene product latent membrane protein 1 engages the tumor necrosis factor receptor-associated death domain protein to mediate B lymphocyte growth transformation and activate NF-kappaB. Proc Natl Acad Sci U S A 1997;94:12592–7.

(36) Devergne O, Hatzivassilliou E, Izumi KM, Kaye KM, Kleijnen MF, Kieff ED, et al. Association of TRAF1, TRAF2, and TRAF3 with an Epstein-Barr virus LMP1 domain important for B-lymphocyte transformation: role in NF-kappaB activation. Mol Cell Biol 1996;16:7098–108.
(37) Heslop HE, Ng CY, Li C, Smith CA, Loftin SK, Krance RA, et al. Long-term restoration of immunity against Epstein-Barr virus infection by adoptive transfer of gene-modified virus-specific T lymphocytes. Nat Med 1996;2:551–5.
(38) Lacerda JF, Ladanyi M, Louie DC, Fernandez JM, Papadopoulos EB, O'Reilly RJ. Human Epstein-Barr virus (EBV)-specific cytotoxic T lymphocytes home preferentially to and induce selective regressions of autologous EBV-induced B cell lymphoproliferations in xenografted C.B-17 scid/scid mice. J Exp Med 1996;183:1215–28.
(39) Gaidano G, Dalla-Favera R. Molecular pathogenesis of AIDS-related lymphomas. Adv Cancer Res 1995;67:113–53.
(40) Dolcetti R, Frisan T, Sjoberg J, De Campos-Lima PO, Pisa P, De Re V, et al. Identification and characterization of an Epstein-Barr virus specific T-cell response in the pathologic tissue of a patient with Hodgkin's disease. Cancer Res 1995;55:3675–81.
(41) Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 1994;266:1865–9.
(42) Ensoli B, Gendelman R, Markham P, Fiorelli V, Colombini S, Raffeld M, et al. Synergy between basic fibroblast growth factor and HIV-1 Tat protein in induction of Kaposi's sarcoma. Nature 1994;371:674–80.
(43) Beral V, Peterman TA, Berkelman RL, Jaffe HW. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? Lancet 1990;335:123–8.
(44) Moore PS, Chang Y. Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma patients with and without HIV infection. N Engl J Med 1995;332:1181–5.
(45) Staskus KA, Zhong W, Gebhard K, Herndier B, Wang H, Renne R, et al. Kaposi's sarcoma-associated herpesvirus gene expression in endothelial (spindle) tumor cells. J Virol 1997;71:715–9.
(46) Kedes DH, Operskatski E, Busch M, Kohn R, Flood J, Ganem D. The seroepidemiology of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus): distribution of infection in KS risk groups and evidence for sexual transmission. Nat Med 1996;2:918–24.
(47) Gao SJ, Kingsley L, Li M, Zheng W, Parravicini C, Zeigler J, et al. KSHV antibodies among Americans, Italians, and Ugandans with and without Kaposi's sarcoma. Nat Med 1996;2:925–8.
(48) Cesarman E, Chang Y, Moore PS, Said JW, Knowles DM. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. N Engl J Med 1995;332:1186–91.
(49) Rainbow L, Platt GM, Simpson GR, Sarid R, Gao SJ, Stobier H, et al. The 222- to 234-kilodalton latent nuclear protein (LNA) of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) is encoded by orf 73 and is a component of the latency-associated nuclear antigen. J Virol 1997;71:5915–21.
(50) Renne R, Zhong W, Herndier B, McGrath M, Abbey N, Kedes D, et al. Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. Nat Med 1996;2:342–6.
(51) Russo JJ, Bohenzky RA, Chien MC, Chen J, Yan M, Maddalena D, et al. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). Proc Natl Acad Sci U S A 1996;93:14862–7.
(52) Neipel F, Albrecht JC, Fleckenstein B. Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? J Virol 1997;71:4187–92.
(53) Jung JU, Desrosiers RC. Association of the viral oncoprotein STP-C488 with cellular ras. Mol Cell Biol 1995;15:6506–12.
(54) Guo J, Duboise M, Lee H, Li M, Choi JK, Rosenzweig M, et al. Enhanced down-regulation of Lck-mediated signal transduction by a Y114 mutation of herpesvirus Saimiri tip. J Virol 1997;71:7092–6.
(55) Miles SA, Rezai AR, Salazar-Gonzalez JF, Vander Mayden M, Stevens RH, Logan DM, et al. AIDS Kaposi sarcoma-derived cells produce and respond to interleukin 6. Proc Natl Acad Sci U S A 1990;87:4068–72.
(56) Moore PS, Boshoff C, Weiss RA, Chang Y. Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV. Science 1996;274:1739–44.
(57) Thome M, Schneider P, Hofmann K, Fickenscher H, Meinl E, Neipel F, et al. Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. Nature 1997;386:517–21.
(58) Sarid R, Sato T, Bohenzky RA, Russo JJ, Chang Y. Kaposi's sarcoma-associated herpesvirus encodes a functional bcl-2 homologue. Nat Med 1997;3:293–8.
(59) McGrath MS, Herndier B, Shiramizu B. Evidence for direct HIV involvement in a broad spectrum of AIDS associated lymphomas. Blood 1994;84:520a.

Note

Supported by Public Health Service grant (CA47006) from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

Human Papillomavirus Infection and Anogenital Neoplasia in Human Immunodeficiency Virus-Positive Men and Women

*Joel M. Palefsky**

Human immunodeficiency virus (HIV)-positive women have a higher prevalence of human papillomavirus (HPV) infection in the cervix and anus, as well as squamous intraepithelial lesions (SILs) at these sites, than do HIV-negative women matched for age and HIV risk factors. Similarly, HIV-positive homosexual or bisexual men have a higher prevalence of anal HPV infection and anal SIL than do HIV-negative homosexual or bisexual men. In HIV-positive individuals, the prevalence of HPV infection, the proportion infected with multiple HPV types, and the prevalence of anogenital SILs increase with decreasing CD4 count. This situation may reflect loss of systemic immune response to HPV antigens or local HPV–HIV interactions at the tissue or cellular level. Despite the high levels of anogenital SILs, to date, there has not been a significant increase in reported cases of invasive anogenital cancer in HIV-positive individuals. However, several years may be required for SIL to progress to invasive cancer, and the advent of newer therapies for HIV that are expected to prolong survival may paradoxically increase the risk of progression to cancer in individuals with SILs if these lesions do not regress spontaneously and remain untreated. [Monogr Natl Cancer Inst 1998;23:15–20]

The association between human papillomavirus (HPV) and invasive cervical cancer has been recognized for many years, initially through the recognition that cervical cancer had the characteristics of a sexually transmitted disease, i.e., association with number of sexual partners and the age at first intercourse. In the last 10 years, with the advent of molecular probes for HPV, a clear association between the presence of HPV DNA in cervical cancer cells has been shown. Furthermore, increasing understanding of the function of HPV proteins known to be expressed in cervical cancer tissues, such as E6 and E7, has established strong biologic plausibility for HPV as a key factor in the pathogenesis of this disease.

Most of current understanding of the epidemiology of HPV infection and anogenital neoplasia was derived from human immunodeficiency virus (HIV)-negative study populations. In recent years, it has become apparent that the prevalence of anogenital HPV infection is higher among HIV-positive men and women than in their HIV-negative counterparts. Likewise, potentially precancerous lesions of the anogenital region are more common among HIV-positive men and women. While this increased prevalence has not yet led to a substantial increase in the incidence of anogenital cancer in these populations, the incidence of cancer may well increase as HIV-positive patients live longer as a result of improvements in medical therapy for HIV infection. The consequences of HPV infection may be a paradigm for a possible shift in complications of HIV disease from acute opportunistic infections to malignancies associated with chronic viral infection. An understanding of HPV infection and anogenital neoplasia in HIV-positive men and women is therefore important, particularly since many, if not all, of the HPV-associated anogenital cancers that occur in these patients may be preventable.

The goals of this article are to summarize current understanding of the role of HPV in the pathogenesis of anogenital neoplasia, to describe current knowledge of the epidemiology of HPV infection and anogenital neoplasia in HIV-positive men and women, to describe HIV–HPV interactions that may play a role in anogenital disease pathogenesis in HIV-positive individuals, and to speculate on the effect of improved therapy for HIV infection on the natural history of anogenital neoplasia.

Role of HPV in Pathogenesis of Anogenital Neoplasia

There are many different anogenital HPV types, and these are generally divided into oncogenic and nononcogenic types by virtue of the frequency of their association with invasive cervical cancer. Of the oncogenic types, HPV type 16 is the most important by virtue of its being the most common in cervical cancer, followed by HPV types 18, 31, and 45 *(1)*. Many HPV types are also found, although with less frequency, in some cases of cervical cancer; these include HPV types 33, 35, 39, 52, 56, 58, 59, and 68. Lastly, there is a large group of nononcogenic types rarely if ever found by itself in cervical cancer, and the ones that are most prominent in the genital region are types 6, 11, 42, 43, and 44.

The mechanisms of oncogenicity of HPV are increasingly, but not completely, understood. The HPV E6 protein binds and degrades its cellular p53 protein target through a ubiquitin-mediated pathway, whereas the HPV E7 protein inactivates the cellular RB protein, as reviewed previously *(2)*. Among their many functions, p53 and RB shut down the cell cycle and negatively regulate cell growth. The p53 protein also has been shown to stimulate DNA repair enzymes after DNA damage, presum-

**Correspondence to:* Joel M. Palefsky, M.D., Department of Laboratory Medicine and Stomatology, Box 0100, Rm. C634, University of California, San Francisco, San Francisco, CA 94143.

See "Notes" following "References."

© Oxford University Press

ably to limit the amount of chromosomal damage that occurs in a cell *(3,4)*. Consequently, through the effects of E6 and E7, HPV infection may lead to genomic instability in cells that continue to cycle despite the presence of chromosomal damage *(3,5)*. Consistent with this hypothesis, chromosomal mutations as shown by studies of loss of heterozygosity may be found in cervical cancer *(6–8)*. The stage at which these changes occur is not yet known, but one or more genetic changes may be required for progression from high-grade squamous intraepithelial lesion (HSIL) to cancer.

Cervical HPV Infection and Cervical Squamous Intraepithelial Lesions (SILs)

HPV on the cervix typically infects in the transformation zone, where the squamous epithelium of the exocervix meets the columnar epithelium of the endocervix. This squamocolumnar junction is a relatively thin, highly metabolically active area of epithelium. Most HPV infections occur here, and most HPV-related lesions, including invasive cancer, arise from this area. A spectrum of histopathologic abnormalities resulting from HPV infection has been described. At the more benign end of the disease spectrum are changes described as low-grade SILs (LSILs). The primary features of LSILs include limited proliferation of basal or parabasal cells with a high nuclear-to-cytoplasmic ratio and koilocytosis (characterized by cells in the more differentiated cell layers with an enlarged, irregular nucleus surrounded by a clear area or halo). At the other end of the disease spectrum are changes described as HSILs. The cardinal feature of HSILs is marked proliferation of immature basal cells, whieh may replace most or all of the normal epithelium, and mitoses in the more superficial cell layers. The clinical importance of this grading scheme is that almost all the invasive cancers arise from HSILs and few, if any, arise from LSILs. Consistent with this observation, almost all HSILs contain high- or medium-risk oncogenic HPV types, whereas LSILs may contain a wider range of types *(9)*.

Epidemiologic studies of cervical HPV infection suggest that the age-related prevalence of HPV infection as determined by polymerase chain reaction (PCR) is highest among women in their late teens and early twenties *(10)*. These data suggest that most women acquire HPV infection relatively early after initiation of sexual activity. The age-related prevalence of cervical HPV infection declines thereafter, suggesting that immune mechanisms may either be clearing the HPV infection or reducing it to levels that are undetectable with the use of current technology. In addition, some of this age-related decline may represent a cohort effect, given changes in sexual behaviors that have occurred in the last few decades.

The age-related prevalence of cervical LSILs parallels that of infection, but only a small proportion of women who acquire HPV infection develop clinically detectable LSILs *(10)*. An even smaller proportion of these women develop HSILs or invasive cervical cancer. Currently, the rates in the United States are approximately 8 per 100 000, and many of these cancers are occurring in women who never have Pap smear screening *(11)*. In developing countries, where there is no routine cervical cytology screening, the rates are much higher, and cervical cancer constitutes a major cause of mortality among young women.

Anal HPV Infection and Anal SILs

In the general population, anal cancer is more common among women than among men. In the pre-HIV era, anal cancer was reported in women approximately four times more often than in men, with an incidence of 13 per 1 000 000 per year in the United States *(12)*. In the last 15 years, the incidence of anal cancer has increased over 35% in women (Hauser A: personal communication). In Denmark, rates of anal cancer are lower than those in the United States; however, between 1957 and 1987, the incidence of anal cancer among Danish women more than tripled to 7.4 per 1 000 000 *(13)*. Rates of anal cancer also rose among Danish men during that time, increasing by 1.5-fold to 3.8 per 1 000 000.

The precise incidence of anal cancer among men with a history of receptive anal intercourse has been difficult to determine because cancer registries do not collect information on sexual orientation or behavior. Men with a history of receptive anal intercourse, however, may be at especially high risk of anal cancer. Daling et al. *(14)* estimated that the incidence of anal cancer among homosexual men was approximately 35 per 100 000, rendering the incidence of anal cancer in this group several times higher than current rates of cervical cancer in women in the United States *(11)* and similar to rates of cervical cancer prior to the introduction of routine cervical cytology screening.

Anal cancer shares many biologic properties with cervical cancer. These cancers are similar histopathologically, and both are strongly associated with HPV infection *(15,16)*. The anal canal has a transformation zone, where the columnar epithelium of the rectum joins the squamous epithelium of the anus, which closely resembles that of the cervix. Like the cervical transformation zone, the anorectal junction is a common site of anal HPV infection and HPV-associated SIL. The histology of anal SIL is similar to that of cervical SIL, and anal HSIL is associated with the same HPV types as cervical HSIL *(15)*. Although studies have never been done to determine if untreated anal HSIL is the precursor to invasive anal cancer, this is likely to be the case, based on knowledge of the natural history of cervical HSIL. From a clinical standpoint, this is important because the anorectal junction is about 2 cm inside the anal canal and thus would not be visible with routine perianal inspection.

Anogenital HPV Infection and SILs in HIV-Positive Men and Women

The data collected on the epidemiology and natural history of HPV infection and SIL described above were derived from HIV-negative individuals. There are several reasons to suspect that these data may be different in HIV-positive men and women. First, both HIV and HPV are sexually transmissible. Behaviors that increase the risk of acquiring HIV infection may also increase the risk of acquiring HPV infection and vice versa. Second, since the immune response to HPV may play an important role in control of SIL, it is possible that, as HIV-related immunosuppression increases, immunity specific to HPV declines. Consequently, HIV-positive individuals may be at higher risk of both acquiring HPV and developing SIL related to HPV once HPV is acquired.

Clear patterns are emerging from studies of large numbers of

HIV-positive men and women. These patterns confirm the increased risk of HPV infection and SIL in these groups. One of these studies is a prospective cohort study known as the WIHS Study, the Women's Interagency HIV Study. This ongoing study enrolled 2015 HIV-positive women and 577 HIV-negative women who were matched for age and HIV risk factors at six different cities around the United States. HPV testing of a cervicovaginal lavage specimen was performed on the women at baseline by use of PCR. The procedure used was a modification of that employed by Ting et al. *(17)* and used consensus primers from the L1 region of the HPV genome, followed by specific typing for 39 different HPV types. Among these high-risk, HIV-negative women, the prevalence of HPV was approximately 26% (Palefsky JM: unpublished data), relatively high for their age when compared with the data of Schiffman *(10)*, which were collected in a population of women with Kaiser Permanente medical insurance. However, the HIV-positive women had even higher rates of prevalent HPV infection, and the rates were highest among those with the lowest baseline CD4 count. Among women with CD4 levels less than 200/mm^3 at baseline, approximately 70% had detectable HPV infection (Palefsky JM: unpublished data). Independent risk factors for cervical HPV infection at baseline included HIV status, current smoker, younger age, and ethnicity (with African-American women having the highest rates of HPV infection).

Similar data have been obtained from a cohort study of 346 HIV-positive and 262 HIV-negative homosexual or bisexual men in San Francisco. By use of a PCR HPV testing method similar to that used for cervicovaginal lavage specimens in the WIHS study, anal HPV infection was characterized in this study population at baseline. Approximately 60% of the HIV-negative men had anal HPV infection. Similar to our findings in the cervix, the proportion of men with anal HPV infection was even higher among the HIV-positive group and was highest among those with the lowest CD4 levels. Among the HIV-positive men with CD4 counts less than 500/mm^3, HPV infection was nearly universal (Palefsky JM: unpublished data). These data were similar to those reported previously among homosexual men in Seattle *(18)*.

Another interesting feature of HPV infection among both HIV-positive women and men was the multiplicity of HPV types. Of HIV-negative women in the WIHS study who were HPV-positive, 16% had more than one HPV type, compared with 42% of HIV-positive women (Palefsky JM: unpublished data). Among HIV-negative men, 23% had more than one type of HPV in the anal canal, compared with 73% of HIV-positive men (Palefsky JM: unpublished data).

With respect to cervical cytologic changes in the WIHS cohort at baseline, a relatively high proportion of HIV-negative women had abnormal cervical cytology (16%). Similar to the HPV data, the proportion of HIV-positive women with abnormal cervical cytology at baseline increased inversely with the baseline CD4 levels, and 53% of women with CD4 levels less than 200/mm^3 had abnormal cytology (Fruchter R: personal communication). Finally, among HIV-negative men in the San Francisco cohort study, 21% had anal cytologic abnormalities at baseline, while 72% of HIV-positive men with CD4 counts less than 200/mm^3 had abnormal anal cytology (Palefsky JM: unpublished data). Taken together, these findings indicate that HIV-positive women and men have a higher prevalence of anogenital HPV infection, a higher number of HPV types, and a higher rate of prevalent anogenital lesions—each of these clearly associated with lower CD4 levels.

Relatively few data are currently available on the natural history of HPV infection and SIL in these large study groups, and these studies are in progress. However, some prospective data on the projected incidence of anal HSILs are available from the San Francisco men's cohort study. Men who entered the study without HSIL were stratified according to HIV status and baseline CD4 level. Among HIV-positive men who had CD4 levels less than 200/mm^3 at baseline or between 200/mm^3 and 500/mm^3, the 4-year projected incidence of anal HSIL was greater than 50% (Palefsky JM: unpublished data). Even among those who entered the study with a CD4 level greater than 500/mm^3, the projected incidence of HSILs was high (approximately 30%). Finally, HIV-negative men in the study had a projected 4-year incidence of 17%, putting them potentially at the highest risk, since many of these men will have a completely normal life span. These data indicate that a high proportion of HIV-positive men, as well as a substantial proportion of HIV-negative men, will develop HSILs if they are followed for a sufficient period of time.

Invasive Anogenital Cancer in HIV-Positive Women and Men

The above data indicate a high prevalence and incidence of anogenital HPV infection and putative precancerous anogenital lesions. An important question that has emerged in the last few years is whether these findings will translate into higher rates of invasive anogenital cancer. Thus far, there has been no significant increase in the rate of invasive cervical cancer among HIV-positive women in the United States or in the developing world *(19)*. Data on anal cancer are not as clear, since studies using different methods lead to different conclusions. Rabkin and Yellin *(19)* did not find a significant increase in anal cancer in single, never-married men in the San Francisco Bay Area in the early 1990s when compared with the pre-HIV years. In contrast, Melbye et al. *(20)* reported increased relative risk of anal cancer compared with the general population with increasing proximity to an AIDS diagnosis, implying a role for increasing immunosuppression. However, these data should be interpreted cautiously, since the results may be confounded by the length of time that an individual may have been infected with HPV. Overall, it seems likely that there has been some increase (but not very large) in anal cancer, and the exact magnitude is still unknown.

Based on knowledge of the natural history of cervical disease, it seems likely that progression of HSIL to invasive cancer may require several years. Until recently, most HIV-positive individuals would have died of other HIV-related complications before HSIL would have had the chance to progress, which may explain why the high prevalence of anogenital HSIL in HIV-positive women and men has not led thus far to a dramatically large increase in cervical or anal cancer.

Despite the absence of a clear increase in cervical cancer in HIV-positive women, it is important to recognize that, once it develops, cervical cancer may be very aggressive in HIV-positive women, and extreme vigilance on the part of the clini-

cian is very important *(21,22)*. Conversely, cervical cancer may be an indicator of HIV infection, particularly in those regions with the highest HIV prevalence, and women with cervical SIL or cervical cancer should be strongly considered for HIV testing.

In the case of cervical disease, the clinical implications of these findings are that aggressive screening, treatment, and follow-up for cervical SIL are necessary. In the case of anal disease, the data indicating the need for an anal-screening program modeled on that used for cervieal disease are compelling. Studies have been performed recently that validate the use of anal colposcopy as a diagnostic tool *(23)* and anal cytology as a screening tool *(24)*. Barriers to implementation of such a screening program include the paucity of studies showing that anal HSIL progresses to invasive cancer and that a screening program would have an impact on the rate of anal cancer. However, such studies are very unlikely to be performed, given the ethical issues of following an individual without treatment once HSIL is diagnosed. Other barriers include the absence of widespread expertise in the performance of anal diagnostic techniques and the morbidity associated with most treatments for anal lesions. Were such a screening program to be implemented in the future, individuals most likely to benefit would include HIV-positive and HIV-negative men with a history of receptive anal intercourse. Because of data showing that women with cervical HSIL or vulvar cancer have a high prevalence of anal SIL and anal cancer *(25,26)*, HIV-positive and HIV-negative women with high-grade eervical or vulvar lesions should also be considered for screening. Data on anal HPV infection and anal SIL in HIV-positive women without cervical disease are still emerging, but several earlier studies *(27,28)* have shown that, similar to cervical HPV infection and cervical SIL, the prevalence of anal lesions is highest among HIV-positive women with the lowest CD4 levels.

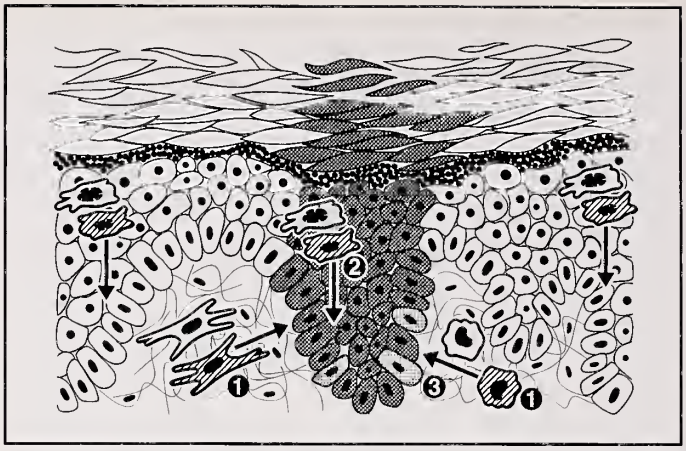


Fig. 1. Schematic representation of possible interactions between human immunodeficiency virus (HIV) and human papillomavirus (HPV) at the tissue and cellular levels in the pathogenesis of anogenital squamous intraepithelial lesions and invasive anogenital cancer. HPV infection is restricted to the epithelium, as shown in the shaded cells. HIV infection, as shown in the striped cells, is located primarily in stromal cells (**1**) such as circulating T cells and possibly fibroblasts or other stromal elements or Langerhans' cells within the epithelium (**2**). HIV and HPV co-infection in the same epithelial cells (**3**) (dotted cells) is also a theoretical possibility. Possible interactions include secretion of HIV-1 tat by Langerhans' or stromal cells which may up-regulate HPV gene expression or other factors such as cytokines. Cells infected with both viruses may also transactivate each other.

Mechanisms of HIV–HPV Interaction

Several mechanisms of interaction between HIV and HPV may play a role in the higher prevalence and incidence of anogenital SIL in HIV-positive women and men. One such mechanism is the systemic immune response to HPV. It is known that women who have iatrogenic immunosuppression secondary to organ transplantation are at increased risk of developing cervical and vulvar cancers when compared with age-matched women with normal immunity *(29)*. Recently, one study *(30)* showed that a smaller proportion of women with HPV type 16 infection and cervical SIL have cytotoxic T-cell responses to the HPV type 16 E6 and E7 proteins than women who were HPV type 16 positive without cervical disease. Other studies of T-cell proliferative responses *(31,32)* also confirm that HPV-positive women without disease have higher response rates than women with lesions. These data suggest that, once a woman acquires HPV infection, having a cell-mediated immunity (CMI) response is associated with protection against lesion development. Conversely, if HIV infection leads to global impairment of CMI responses, including those to HPV antigens, then loss of that CMI response to HPV antigens may be important in the pathogenesis of anogenital disease and may explain in part the higher levels of HPV infection and SIL among those HIV-positive individuals with the lowest CD4 levels.

Local interactions at the tissue and cellular levels between HIV and HPV may also play a role in potentiation of anogenital neoplasia in HIV-positive individuals (Fig. 1). Although the vi-

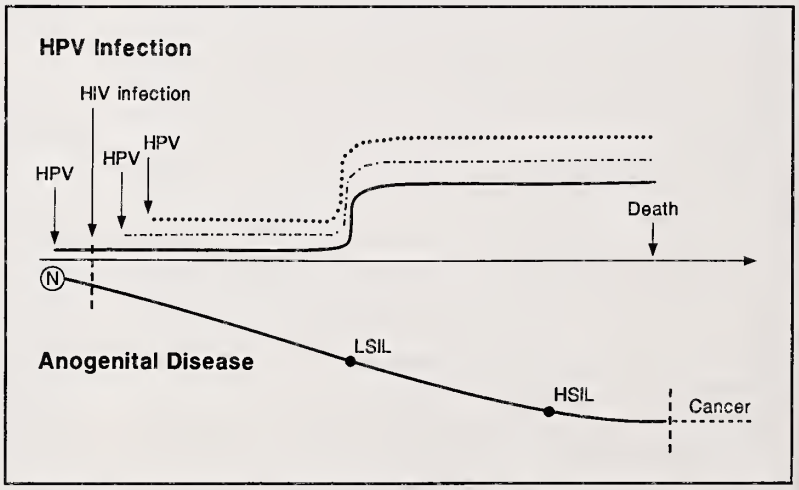


Fig. 2. Model for the role of systemic immune response in the pathogenesis of anogenital squamous intraepithelial lesions (SILs) and invasive anogenital cancer. Human papillomavirus (HPV) infection is acquired early after initiation of sexual activity, followed thereafter by human immunodeficiency virus (HIV) acquisition. Additional HPV types, represented by different lines, may be acquired with new exposures. Early in the natural history of HIV infection, systemic immune responses remain relatively intact. HPV replication levels remain low, and there are no anogenital lesions. With advancing HIV disease, systemic immunity is attenuated, and loss of control of HIV replication is seen as depicted by increased HPV levels. This is accompanied by development of low-grade SIL (LSIL), which may progress over time to high-grade SIL (HSIL) if the lesion remains untreated and the HIV disease continues to progress. In most individuals, progression of HSIL to invasive cancer does not occur because they die of other HIV-related complications first. N = normal.

ruses are likely to be found in different cell types *(33)* (HPV in the epithelium and HIV in local Langerhans' cells and stromal cells), one possibility for interaction between these two viruses locally is through production of HIV-1 tat, which has been shown to be secreted by HIV-infected cells and may be able to up-regulate expression of the HPV type 16 E6 and E7 genes *(34)*. Aberrant cytokine expression by HIV-infected cells may also play a role in potentiating HPV infection. Perhaps most controversial is the possibility of co-infection with HPV and HIV in the same epithelial cell and the possibility of mutual transactivation between the viruses.

Taken together, these disparate observations may be used to construct a model for the interplay between HPV and HIV in the pathogenesis of anogenital SIL and cancer (Fig. 2). In this model, individuals acquire HPV infection relatively early on after initiation of sexual activity and, in many cases, before HIV infection. HIV is acquired at one or more moments in time, and these individuals may continue to acquire different strains of HPV as well. As long as the systemic and local immune responses remain intact, HPV replication and gene expression are controlled and there are no anogenital lesions. If HIV infection is allowed to persist and the immune response deteriorates, higher levels of viral replication and expression of HPV-transforming genes may lead to development of a low-grade lesion. With time, which may vary considerably from one individual to another, the disease may progress to HSIL. As described earlier, the high prevalence of HSIL does not appear to be accompanied by increases in the rate of anogenital cancer because of HIV-related mortality before progression of HSIL to cancer occurs.

Conclusions: What Does the Future Hold?

The natural history of anogenital HPV infection and SIL in the era of highly active antiretroviral therapy (HAART) may represent a paradigm for HIV-related complications of the future. HAART has been shown to dramatically reduce systemic HIV viral load and to reduce the incidence of some opportunistic infections. Because HAART has not been in use for an extended period of time, it is not yet clear how much it will prolong the survival of individuals with HIV infection. Two scenarios are possible (Fig. 3). If individuals go on HAART after developing HSIL and if partial or complete restoration of immune response to HPV occurs, then one might expect HPV levels to decrease and HSIL to regress. If that is the case, then a decreased incidence of anogenital cancer from current levels would be expected. Conversely, if HAART permits individuals to live longer but does not have a significant impact on the immune response to HPV, then the disease in individuals with HSIL might now have the time to progress to invasive cancer, resulting in an increased incidence of anogenital cancer. It is not yet clear which of these two scenarios is correct. However, the most that HAART therapy could be expected to achieve would be to restore the immune system of an HIV-positive individual to one as functional as that of an HIV-negative individual. Since most HSILs do not regress spontaneously, even in HIV-negative individuals, the second scenario seems more likely at this time. Clearly, further natural history studies of men and women on HAART are needed, as are studies of the interactions between HIV and HPV in the pathogenesis of anogenital cancer.

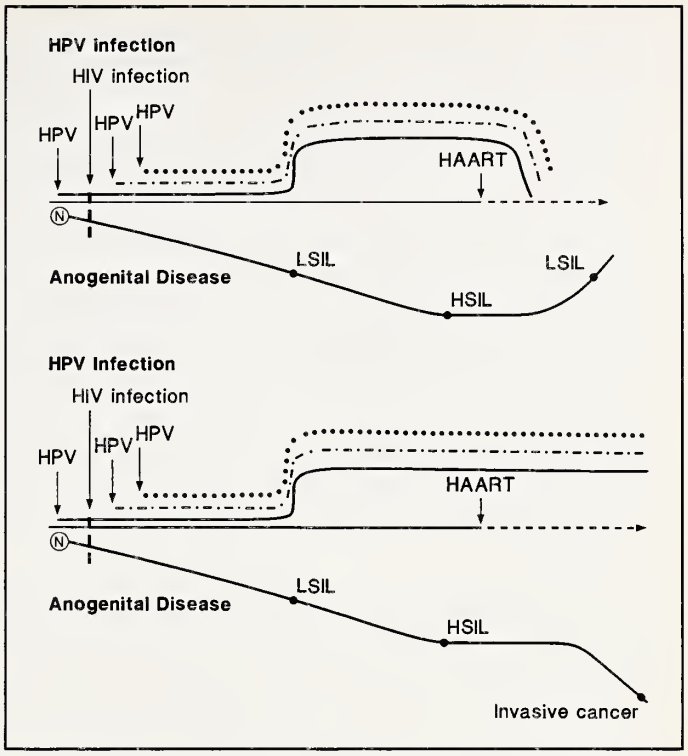


Fig. 3. Model for the role of systemic immune response in the pathogenesis of anogenital squamous intraepithelial lesions (SILs) and invasive anogenital cancer in the era of highly active antiretroviral therapy (HAART). HSIL = high-grade SIL; LSIL = low-grade SIL. **Upper panel:** If HAART therapy results in both restoration of systemic immune responses, including those to human papillomavirus (HPV) antigens, and prolongation of life, then individuals with HSIL may have decreased levels of HPV and regression of HSIL. The net result would be a rate of invasive anogenital cancer that parallels that of human immunodeficiency virus-negative individuals. **Lower panel:** If HAART leads to prolongation of life but does not fully restore immunity to HPV, then no regression of HSIL will be seen. In this case, the prolongation of life may increase the risk of progression to invasive cancer, since there may now be sufficient time for progression before the individual dies of other causes. N = normal.

References

(1) Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International Biological Study on Cervical Cancer (IBSCC) Study Group. J Natl Cancer Inst 1995;87:796–802.

(2) Palefsky JM, Holly EA. Molecular virology and epidemiology of human papillomavirus and cervical cancer. Cancer Epidemiol Biomarkers Prev 1995;4:415–28.

(3) Livingstone LR, White A, Sprouse J, Livanos E, Jacks T, Tlsty TD. Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53. Cell 1992;70:923–35.

(4) Vousden KH, Vojtesek B, Fisher C, Lane D. HPV-16 E7 or adenovirus E1A can overcome the growth arrest of cells immortalized with a temperature-sensitive p53. Oncogene 1993;8:1697–702.

(5) Carder P, Wyllie AH, Purdie CA, Morris RG, White S, Piris J, et al. Stabilised p53 facilitates aneuploid clonal divergence in colorectal cancer. Oncogene 1993;8:1397–401.

(6) Magnusson PK, Wilander E, Gyllensten U. Analysis of loss of heterozygosity in microdissected tumor cells from cervical carcinoma using fluorescent dUTP labeling of PCR products. Biotechniques 1996;21:844–7.

(7) Rader JS, Kamarasova T, Huettner PC, Li L, Li Y, Gerhard DS. Allelotyping of all chromosomal arms in invasive cervical cancer. Oncogene 1996;13:2737–41.

(8) Douc-Rasy S, Barrois M, Fogel S, Ahomadegbe JC, Stehelin D, Coll J, et al. High incidence of loss of heterozygosity and abnormal imprinting of H19 and IGF2 genes in invasive cervical carcinomas. Uncoupling of H19 and IGF2 expression and biallelic hypomethylation of H19. Oncogene 1996;12:423–30.

(9) Lorincz AT, Reid R, Jenson AB, Greenberg MD, Lancaster W, Kurman RJ. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. Obstet Gynecol 1992;79:328–37.
(10) Schiffman MH. Recent progress in defining the epidemiology of human papillomavirus infection and cervical neoplasia. J Natl Cancer Inst 1992; 84:394–8.
(11) Qualters JR, Lee NC, Smith RA, Aubert RE. Breast and cervical cancer surveillance, United States, 1973–1987. MMWR Morb Mortal Wkly Rep 1992;41:1–15.
(12) Young J Jr, Percy CL, Asire AJ, Berg JW, Cusano MM, Gloeckler LA, et al. Cancer incidence and mortality in the United States, 1973–77. Monogr Natl Cancer Inst 1981;57:1–187.
(13) Frisch M, Melbye M, Moller H. Trends in incidence of anal cancer in Denmark. BMJ 1993;306:419–22.
(14) Daling JR, Weiss NS, Hislop TG, Maden C, Coates RJ, Sherman KJ, et al. Sexual practices, sexually transmitted diseases, and the incidence of anal cancer. N Engl J Med 1987;317:973–7.
(15) Palefsky JM, Holly EA, Gonzales J, Berline J, Ahn DK, Greenspan JS. Detection of human papillomavirus DNA in anal intraepithelial neoplasia and anal cancer. Cancer Res 1991;51:1014–9.
(16) Zaki SR, Judd R, Coffield LM, Greer P, Rolston F, Evatt BL. Human papillomavirus infection and anal carcinoma. Retrospective analysis by *in situ* hybridization and the polymerase chain reaction. Am J Pathol 1992; 140:1345–55.
(17) Ting Y, Manos MM. Detection and typing of genital human papillomaviruses. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR protocols and applications. San Diego (CA): Academic Press, 1990: 356–67.
(18) Kiviat NB, Critchlow CW, Holmes KK, Kuypers J, Sayer J, Dunphy C, et al. Association of anal dysplasia and human papillomavirus with immunosuppression and HIV infection among homosexual men. AIDS 1993;7: 43–9.
(19) Rabkin CS, Yellin F. Cancer incidence in a population with a high prevalence of infection with human immunodeficiency virus type 1. J Natl Cancer Inst 1994;86:1711–6.
(20) Melbye M, Cote TR, Kessler L, Gail M, Biggar RJ. High incidence of anal cancer among AIDS patients. The AIDS/Cancer Working Group. Lancet 1994;343:636–9.
(21) Maiman M, Fruchter RG, Serur E, Remy JC, Feuer G, Boyce J. Human immunodeficiency virus infection and cervical neoplasia. Gynecol Oncol 1990;38:377–82.
(22) Rellihan MA, Dooley DP, Burke TW, Berkland ME, Longfield RN. Rapidly progressing cervical cancer in a patient with human immunodeficiency virus infection. Gynecol Oncol 1990;36:435–8.
(23) Jay N, Holly EA, Berry M, Hogeboom CJ, Darragh JM, Palefsky JM. Colposcopic correlates of anal squamous intraepithelial lesions. Dis Col Rectum 1997;40:919–28.
(24) Palefsky JM, Holly EA, Hogeboom CJ, Jay N, Berry M, Darragh TM. Anal cytology as a screening tool for anal squamous intraepithelial lesions. J Acquir Immun Defic Syndr 1997;14:415–22.
(25) Scholefield JH, Hickson WG, Smith JH, Rogers K, Sharp F. Anal intraepithelial neoplasia: part of a multifocal disease process. Lancet 1992;340: 1271–3.
(26) Ogunbiyi OA, Scholefield JH, Robertson G, Smith JH, Sharp F, Rogers K. Anal human papillomavirus infection and squamous neoplasia in patients with invasive vulvar cancer. Obstet Gynecol 1994;83:212–6.
(27) Williams AB, Darragh TM, Vranizan K, Ochia C, Moss AR, Palefsky JM. Anal and cervical human papillomavirus infection and risk of anal and cervical epithelial abnormalities in human immunodeficiency virus-infected women. Obstet Gynecol 1994;83:205–11.
(28) Melbye M, Smith E, Wohlfahrt Nielson J, Osterlind A, Orholm M, Bergmann OJ, et al. Anal and cervical abnormality in women—prediction by HPV tests. Int J Cancer 1996;68:559–64.
(29) Penn I. Cancer in the immunosuppressed organ recipient. Transplant Proc 1991;23:1771–2.
(30) Nakagawa M, Stites DP, Farhat S, Sisler JR, Moss B, Kong F, et al. Cytotoxic T lymphocyte responses to E6 and E7 proteins of human papillomavirus type 16: relationship to cervical intraepithelial neoplasia. J Infect Dis 1997;175:927–31.
(31) Nakagawa M, Stites D, Farhat S, Judd A, Moscicki AB, Canchola AJ, et al. T cell response to human papillomavirus type 16: relationship to cervical intraepithelial neoplasia. Clin Diagnost Lab Immunol 1996;3:205–10.
(32) Tsukui T, Hildesheim A, Schiffman MH, Lucci JR, Contois D, Lawler P, et al. Interleukin 2 production *in vitro* by peripheral lymphocytes in response to human papillomavirus-derived peptides: correlation with cervical pathology. Cancer Res 1996;56:3967–74.
(33) Nuovo GJ, Forde A, MacConnell P, Fahrenwald R. *In situ* detection of PCR-amplified HIV-1 nucleic acids and tumor necrosis factor. Am J Pathol 1993;143:40–8.
(34) Vernon SD, Hart CE, Reeves WC, Icenogle JP. The HIV-1 tat protein enhances E2-dependent human papillomavirus 16. Virus Res 1993;27: 133–45.

Notes

Supported by Public Health Service (PHS) grants CA54053 and CA63933 from the National Cancer Institute, National Institutes of Health (NIH), Department of Health and Human Services, as well as by PHS grant AI-DE34989 from the National Institute of Allergy and Infectious Diseases and the National Institute of Dental Research, NIH.

We acknowledge the subjects and investigators of the University of California, San Francisco, Anal Cancer Study and the Women's Interagency HIV Study.

Acquired Immunodeficiency Syndrome-Related Cancers: the Community Perspective

*Michael Marco**

Those individuals who work for acquired immunodeficiency syndrome (AIDS) community-based organizations (treatment education, dissemination, advocacy, and care) and the people with AIDS whom they represent are commonly referred to as the "AIDS community." While the primary interest of the AIDS community is safe and effeetive antiretroviral therapy, many of us are concerned with the opportunistic neoplasms that affect approximately 25% of people with AIDS. The community's perspective on AIDS-related cancers can be summarized as follows: 1) It's about time the National Cancer Institute (NCI) got its act together. 2) How and when are diseases such as AIDS-related Kaposi's sarcoma and non-Hodgkin's lymphoma going to be cured and prevented? 3) Until then, how can we be sure that we will receive the best medical care for these cancers and the underlying human immunodeficiency virus (HIV) infection?

With regard to NCI's finally getting its act together, my perception is that the AIDS community is pleased that there has been a change in leadership at the NCI. The previous era gave us zidovudine (AZT) and didanosine, but unfortunately little was done when it came to AIDS-related cancers. Many of us have wondered why the NCI—with its 250 million dollars in AIDS research funds—is doing so little extramurally to address these cancers. There are some in the AIDS community who trace this failure to the division of labor among the various institutes of the National Institutes of Health (NIH) regarding AIDS.

More recently, the NCI should be praised for spearheading new research initiatives on AIDS-related cancers. The NCI recently formed the AIDS Malignancy Consortium (AMC), a clinical trials network that concentrates solely on AIDS-related cancers. In its first year of operation, the AMC already has five studies open to treat either Kaposi's sarcoma or non-Hodgkin's lymphoma; two of these studies use conventional chemotherapy together with protease inhibitors, and three studies are employing cytokine inhibitors or immune modulators.

The vast majority of the AMC members came directly from the AIDS Clinical Trials Group's (ACTG) Oncology Committee, whose clinical trials in the late 1980s and early 1990s helped develop a clinical standard of care for Kaposi's sarcoma and non-Hodgkin's lymphoma. During that period, we knew very little about the pathogenesis of Kaposi's sarcoma and non-Hodgkin's lymphoma, and it was up to these researchers to determine the proper therapeutic doses of the limited chemotherapeutic agents that we knew had activity. Thus, these clinical trials taught us how to care properly for people with AIDS who were diagnosed as having Kaposi's sarcoma and non-Hodgkin's lymphoma with a scant arsenal of drugs.

The ACTG Oncology Committee was, however, working within a clinical trials network filled with infectious disease specialists who had limited expertise in or commitment to AIDS-related cancer clinical research. The oncologists, who used less than 10% of the ACTG's financial resources, were relegated to a lesser role.

Now that the AMC is in its second year, we should ensure its success by providing it with the resources that it needs. The first resource is, of course, money. Increased funding will help laboratory intensive studies that will measure cytokine levels as well as patient's HIV RNA.

The second resource is patients for its studies. For large phase II and phase III studies that are generated from the AMC or the Eastern Cooperative Oncology Group's AIDS Committee, we will need a majority of the cooperative oncology groups to sponsor these protocols. If this does not work, we might have to return to ACTG for patients.

With regard to advisory panels, the NCI needs to use the members of its AIDS Malignancy Working Group (AMWG) more actively. Many from the AMWG helped organize the excellent (and badly needed) National AIDS Malignancy Conference at the NIH, but putting on a conference is not enough. The NCI should form a working group to discuss and formulate AIDS-related cancer RFAs (i.e., requests for application). Yes, there are some legal matters to consider, since some of these experts in the working group would themselves apply, but we should try to work around this situation. The NCI cannot successfully develop such programs without the full involvement of AMWG. The AMWG should be more than window dressing.

The community's second concern can be summarized in the following question: How and when are cancers, such as Kaposi's sarcoma and non-Hodgkin's lymphoma, going to be cured and prevented?

This is a difficult question, especially coming from the AIDS community who, understandably, has never been known for its patience. The cure for these cancers will eventually come from the work completed by many of the researchers who presented papers at the National AIDS Malignancy Conference. We are just beginning to understand the etiology and pathogenesis of Kaposi's sarcoma and non-Hodgkin's lymphoma. Basic scientists and clinicians must work together in the old-fashioned "bench-to-clinic" mode. When trying to turn their discoveries into active agents, basic scientists must use clinicians as consultants.

For example, there is still much to learn about Kaposi's sarcoma herpesvirus (KSHV). We must first answer three integral

**Correspondence to:* Michael Marco, Treatment Action Group, 105 West 73rd St., Apartment 1D, New York, NY 10023.

© Oxford University Press

questions: 1) Is KSHV a true cause of Kaposi's sarcoma or just a helper virus? 2) What are the roles of cytokines and growth factors? 3) Which assay is most sensitive, and how soon can we start using these assays for screening patients?

In our quest for these cures, industry—especially the biotech companies—will need to take a more active role in AIDS-related cancer clinical research. They must provide their agents to knowledgeable AIDS oncologists for pilot studies or to the NCI for use in the AMC or cooperative oncology group studies.

The U.S. Food and Drug Administration (FDA) must also get more involved. Its Oncologic Drug Division needs to fully understand all aspects of AIDS-related Kaposi's sarcoma and AIDS-related non-Hodgkin's lymphoma. While it should ensure that industry carries out safe and well-designed clinical trials, the Oncologic Drug Division must realize that AIDS-related Kaposi's sarcoma not only causes morbidity and mortality but also emotionally damages patients with AIDS. The Oncologic Drug Division should ensure that the Oncologic Drug Advisory Committee has at least three reviewers who are knowledgeable in all aspects of AIDS-related cancers when AIDS-related Kaposi's sarcoma and AIDS-related non-Hodgkin's lymphoma new drug applications (NDAs) are reviewed. Likewise, the Oncologic Drug Division should attempt to work closely with the Antiviral Drug Division when it is reviewing AIDS-related Kaposi's sarcoma and AIDS-related non-Hodgkin's lymphoma NDAs.

The FDA and the NCI should be applauded for working together on developing clinical end points for Kaposi's sarcoma that will be used alongside classic tumor response. The use of these newly developed clinical end points, such as pain, skin lesion color, and edema, will help us evaluate the true effectiveness of experimental agents.

To test new agents and possibly use these new end points, we will need patients for these studies. Accrual for studies on AIDS-related cancer, like that for most cancer studies, is abysmal. One solution to this problem might be to form close ties with primary care physicians and AIDS clinicians at hospitals and clinic sites so that patients with Kaposi's sarcoma and non-Hodgkin's lymphoma are routinely referred to AIDS oncologists and are informed of studies that have opened.

The final concern of the AIDS community is a frequently asked question: How can we be sure that we will receive the best medical care for these cancers and the underlying HIV infection? To ensure that patients receive the best medical care for their cancers, we must attempt to adopt standard (and evolving) practices and principles of the therapy. As stated earlier, clinical trials that were completed in the last 10 years helped us to understand what drugs (and at what doses) were best for patients with disease at various stages. We know how to use radiation, interferon, ABV (i.e., doxorubicin–bleomycin–vinblastine), liposomal doxorubicin, liposomal daunorubicin, and paclitaxel to treat Kaposi's sarcoma, and we know that modified doses of doxorubicin hydrochloride and cyclophosphamide are warranted for the severely immunocompromised patients with AIDS-related non-Hodgkin's lymphoma.

Although convening a government panel to draft practices and principles of therapy for AIDS-related Kaposi's sarcoma was suggested to the NCI at its AMWG meeting in September 1996, no action has yet been taken on this recommendation. My sense is that many in the AIDS community believe that this lack of action represents an abdication of the government's public health responsibility.

The U.S. Public Health Service convened two panels that recently published guidelines for anti-HIV therapy and treatment for opportunistic infections. Even though the field is still evolving in these areas, the guidelines will ensure that patients receive the best up-to-date AIDS clinical care.

Nevertheless, a Kaposi's sarcoma practices and principles of therapy panel was convened. However, it was sponsored by a small biotech company that took the idea and ran with it. Granted, this company has a financial interest in the outcome of these recommendations, but it was able to assemble an excellent panel of experienced AIDS oncologists, an AIDS clinician, a dermatologist, and an radiation oncologist to discuss methods of treatment of patients with various clinical presentations of Kaposi's sarcoma.

The recommendations that will come from this panel will not hold as much weight as ones that should come from the NCI. It is necessary to give certain guidelines to all those who treat patients with Kaposi's sarcoma. To instruct a physician about when it is time to use systemic therapy with either interferon or cytotoxic chemotherapy and when it is no longer appropriate to use cryotherapy or intralesional vinblastine will undoubtedly help patients.

Until the NCI decides that this a worthwhile initiative, they should see to it that the new Public Health Service guidelines for HIV therapy and opportunistic infection treatment and prophylaxis get into the hands of as many oncologists as possible. For the treatment of patients with Kaposi's sarcoma, non-Hodgkin's lymphoma, or anal or cervical dysplasia, the underlying HIV infection must always be taken into account and treated properly.

The AIDS community is anxious to see new developments in AIDS-related cancer research. Now that people with AIDS are living longer with the advent of protease inhibitors in combination with nucleoside analogues, it is certain that the rates of AIDS-related Kaposi's sarcoma and non-Hodgkin's lymphoma will increase. Thus, clinical and basic AIDS-related cancer research needs to be of concern for all those working in AIDS.

Association of Non-Acquired Immunodeficiency Syndrome-Defining Cancers With Human Immunodeficiency Virus Infection

*Charles S. Rabkin**

Kaposi's sarcoma and non-Hodgkin's lymphoma were among the earliest recognized manifestations of the acquired immunodeficiency syndrome (AIDS) epidemic. Excluding these two tumors, the overall risk of all other cancers in human immunodeficiency virus (HIV)-infected individuals is similar to that of the general population. However, varying levels of evidence link several additional neoplasms to HIV infection. The evidence is strongest for an association with Hodgkin's disease, with lower relative and absolute risks than for non-Hodgkin's lymphoma. Anogenital intraepithelial neoplasia also appears to be HIV associated, but increases of invasive disease are still uncertain for both cervical and anal cancers. Various studies have suggested associations with testicular seminoma, multiple myeloma, oral cancer, and melanoma, but the data are inconsistent. Leiomyosarcoma and benign leiomyomas have increased in incidence in HIV-infected children but are unusual in HIV-infected adults. Conjunctival carcinoma is seen in HIV-infected individuals in sub-Saharan Africa but it is uncommon in Western countries. Most other cancers do not seem to have increased incidences in HIV infection. The etiologic mechanisms of HIV-related cancer likely differ among these diverse cancers and do not globally increase cancer risk. [Monogr Natl Cancer Inst 1998;23:23–25]

Kaposi's sarcoma and non-Hodgkin's lymphoma are relatively frequent outcomes of human immunodeficiency virus (HIV) infection. Several additional tumors appear to be associated with HIV infection, albeit with smaller relative and absolute risks. The evidence is strongest for associations of HIV with anogenital neoplasia, Hodgkin's disease, testicular seminoma, pediatric leiomyosarcoma, and conjunctival cancer. Most other cancers, however, including the carcinomas most common in the general population, do not appear to be increased in HIV infection.

Excluding Kaposi's sarcoma and non-Hodgkin's lymphoma, the overall risk of all other cancers in HIV-infected individuals is similar to that of the general population. The Viral Epidemiology Branch follows two cohorts of HIV-infected hemophilia patients. The cancer experience (through March 1991) of the 1261 HIV-infected subjects in the Multicenter Hemophilia Cohort Study has been previously reported *(1)*. The National Cancer Institute Registry of HIV-Infected Hemophilia Patients follows an additional 1639 subjects recruited in 1991 and 1992. Through June 1996, there were a total of 7675 person-years of observation for the two cohorts combined. With the exclusion of Kaposi's sarcoma and non-Hodgkin's lymphoma, a total of 20 cases of other cancers were observed in comparison with the 13.2 expected, for a relative risk (RR) of 1.5 (95% confidence interval [CI] = 0.9–2.3). Unlike non-Hodgkin's lymphoma incidence, the incidence of other cancers did not significantly increase with the duration of HIV infection (Fig. 1).

HIV and concomitant immunosuppression has been associated with cervical carcinoma *in situ* in studies by Vermund et al. *(2)*, Williams et al. *(3)*, Klein et al. *(4)*, Ho et al. *(5)*, and others. Although cervical cancer was added to the 1993 surveillance definition for AIDS by the U.S. Centers for Disease Control and Prevention, an effect of HIV on invasive cancer is uncertain. Between 1976 and 1988, invasive cervical cancer incidence decreased approximately 40% in New York City black women, a group with a high prevalence of HIV infection *(6)*. Furthermore, New York City AIDS cases from 1992 through 1993 had only a fourfold increase of cervical cancer relative to the general population *(7)*, despite likely confounding by shared behavioral risk factors for HIV and human papillomavirus (HPV) infection. Similarly, there was only one case of invasive cervical cancer in 3612 woman-years of observation after AIDS in the AIDS–Cancer Match Registry (Goedert JJ: personal communication), which matches population-based cancer and AIDS registries in Puerto Rico and seven regions of the United States *(8)*. While the impact of Pap screening is difficult to assess, a large excess risk of invasive cervical cancer seems unlikely by these data.

Anal intraepithelial neoplasia has also been associated with HIV and/or HIV-related immunosuppression in studies by Palefsky et al. *(9)*, Kiviat et al. *(10)*, and others. As with cervical cancer, an etiologic association of HIV with invasive anal cancer is uncertain. Incidence in single San Francisco men aged 25–54 years (a population approximately 50% homosexual) was 2.0 per 100 000 from 1973 through 1979, which was 9.9 (95% CI = 4.5–18.7) times that of the general population in the same period. Thus, even before the HIV epidemic, these men were at very high risk of invasive anal cancer. While the rate increased to 3.9 per 100 000 in 1988 through 1990, incidence in the general population increased in parallel, so the RR was virtually unchanged at 10.1 (95% CI = 5.0–18.0) times expected *(11)*. In more recent data through 1994, the number of anal cancer cases

**Affiliation of author:* Viral Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

Correspondence to: Charles S. Rabkin, M.D., M.Sc., National Institutes of Health, Executive Plaza North, Rm. 434, Bethesda, MD 20892.

See "Note" following "References."

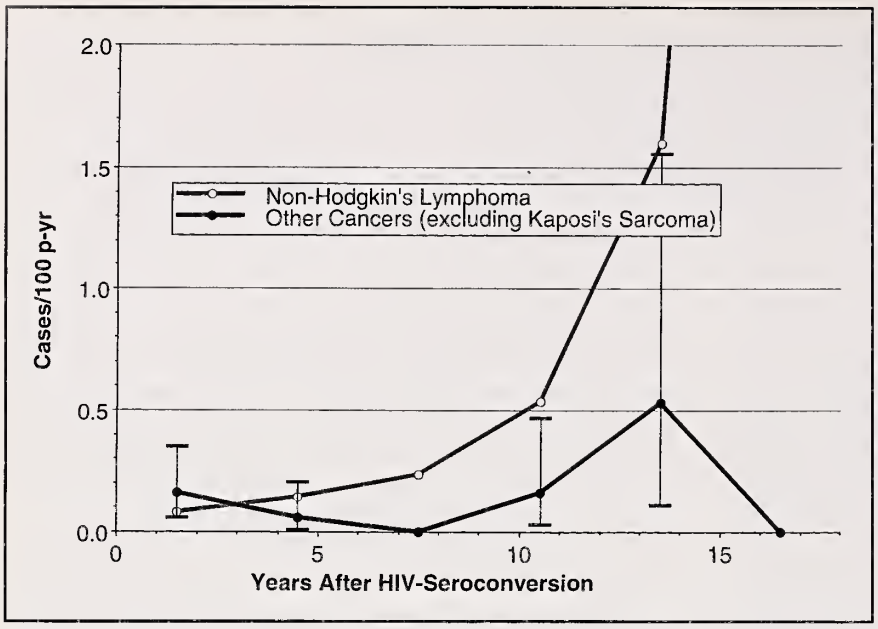


Fig. 1. Incidence of non-Hodgkin's lymphoma and of other eancers (excluding Kaposi's sarcoma) in the Multicenter Hemophilia Cohort Study by duration of HIV infection.

in San Francisco men has continued to increase. Moreover, about half of the cases since 1985 have occurred in association with AIDS, diagnosed either before or after anal cancer (Fig. 2). While these data may reflect some excess risk with advanced HIV infection, the relative excess appears to be small in comparison with the 10-fold RR predating AIDS.

Early evidence that Hodgkin's disease is increased by HIV infection came from the San Francisco City Clinic Cohort *(12)*, which has been updated in a combined analysis (total 15 565 subjects) with the New York City hepatitis B natural history and vaccine trial cohorts *(13)*. Additional evidence came from a San Francisco AIDS and cancer registry-linkage study *(14)*. These findings have been corroborated in several additional studies, including the Multicenter AIDS Cohort Study *(15)* (2683 subjects, 17 780 HIV-positive person-years), the New South Wales AIDS–Cancer Match *(16)* (3616 subjects), and the National Cancer Institute hemophilia studies. The RR of Hodgkin's disease was 2.5 in the hepatitis B cohorts and ranged from 5.6 to 8.5 in the other studies; the 95% CI excluded 1.0 in all four studies (Table 1). The hepatitis cohort found the lowest risk of total non-AIDS cancers (Table 1), perhaps because cases were ascertained by matching with cancer registrations rather than by active follow-up.

The four studies are less consistent with respect to HIV-associated increases in other cancers (Table 1). An increase in testicular seminoma was previously reported in the Multicenter AIDS Cohort Study by Lyter et al. *(15)* but was not observed in the other studies. Statistically significant increases (i.e., 95% CI excluding 1.0) of multiple myeloma and oral cancer were observed in the Multicenter AIDS Cohort Study and the New South Wales AIDS–Cancer Match. Malignant melanoma was significantly increased in the Multicenter AIDS Cohort Study yet not in New South Wales where the disease is more common. These inconsistent associations warrant further investigation before being accepted or discarded.

Notably absent from this list are increases in leiomyosarcoma and squamous cell conjunctival cancer. HIV-infected children are at greatly increased risk of leiomyosarcoma and benign leiomyomas, as first noted by Chadwick et al. *(17)*. These tumors have only rarely been noted in HIV-infected adults *(18)*, despite the much larger number potentially at risk. The HIV-associated cases uniformly contain Epstein-Barr virus (EBV), leading to speculation that prior EBV infection protects HIV-infected adults from this disorder. Similarly, conjunctival squamous cell cancer is found relatively frequently in AIDS cases in sub-Saharan Africa *(19)*, yet is rare in the United States. Differences in solar ultraviolet radiation exposure or in prevalence of conjunctival HPV infection may explain the variation.

Despite their limited variety, HIV-associated tumors are likely to be diverse in their pathology and etiology. Viral cofactors, such as EBV, HPV, and human herpesvirus type 8 may play a role in some of these disorders, but not all viral-associated tumors are increased in AIDS. Notable exceptions are hepatocellular carcinoma (despite high prevalence of both hepatitis B and hepatitis C in HIV-infected hemophilia patients) and nasopharyngeal carcinoma (despite near universal prevalence of EBV). Immunodysregulation or cytokine imbalance may underlie the various tumor associations. More importantly, even advanced immunodeficiency does not appear to lead to in-

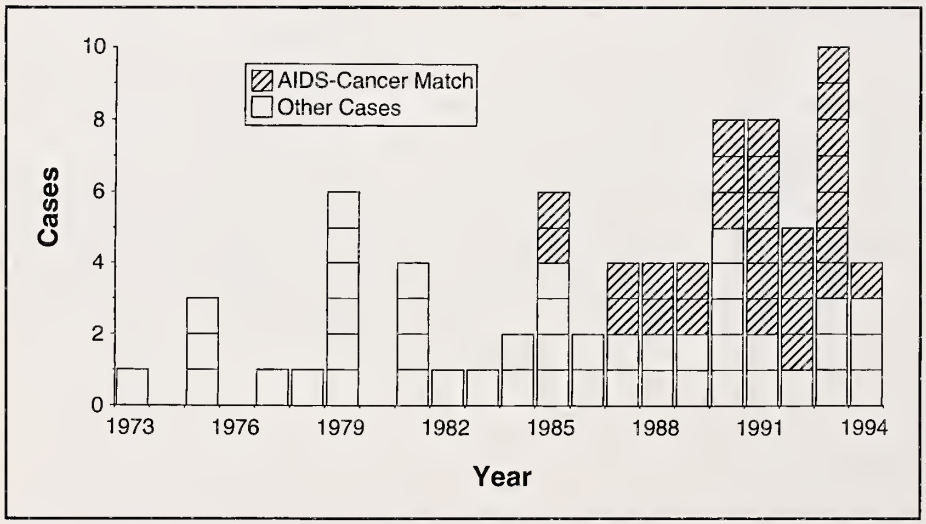


Fig. 2. Anal cancer cases and AIDS comorbidity in San Francisco County men aged 25–54 years, SEER[1] Program and AIDS–Cancer Match Registry, from 1973 through 1994. Shaded boxes = cases with AIDS at or after cancer diagnosis; unshaded boxes = difference in a given year between the total cases detected by cancer surveillance and the number detected by the AIDS–Cancer Match Registry.

Table 1. Relative risks of selected cancers in cohort- and registry-matching studies of HIV and cancer*

Cancer type or site	Cohort	No. of cases	Relative risk	95% CI
Total non-AIDS	NCI Hemophilia Cohort and Registry	20	1.5	0.9–2.3
	Multicenter AIDS Cohort Study	51	2.6	1.9–3.4
	NYC and SF Hepatitis B Studies	168	0.7	0.6–0.8
	New South Wales AIDS–Cancer Match	70	—	—
Hodgkin's disease	NCI Hemophilia Cohort and Registry	3	5.6	1.1–16
	Multicenter AIDS Cohort Study	5	6.7	2.2–16
	NYC and SF Hepatitis B Studies	18	2.5	1.5–3.9
	New South Wales AIDS–Cancer Match	10	8.5	4.1–16
Testicular seminoma	NCI Hemophilia Cohort and Registry	2	2.5	0.3–8.8
	Multicenter AIDS Cohort Study	7	3.9	1.6–8
	NYC and SF Hepatitis B Studies	9	0.5	0.2–0.9
	New South Wales AIDS–Cancer Match	4	1.2	0.3–2.9
Multiple myeloma	NCI Hemophilia Cohort and Registry	0	0	—
	Multicenter AIDS Cohort Study	3	14.2	2.9–41
	NYC and SF Hepatitis B Studies	1	0.5	0–2.8
	New South Wales AIDS–Cancer Match	4	5.8	1.2–17
Melanoma	NCI Hemophilia Cohort and Registry	3	3.7	0.8–11
	Multicenter AIDS Cohort Study	6	2.9	1.1–6.3
	NYC and SF Hepatitis B Studies	11	0.6	0.3–1.1
	New South Wales AIDS–Cancer Match	15	1.3	0.8–2.2
Oral	NCI Hemophilia Cohort and Registry	2	3.3	0.4–12
	Multicenter AIDS Cohort Study	5	3.9	1.3–9.1
	NYC and SF Hepatitis B Studies	8	0.6	0.3–1.2
	New South Wales AIDS–Cancer Match	6	4.1	1.6–9.5

*HIV = human immunodeficiency virus; NCI = National Cancer Institute; NYC = New York City; SF = San Francisco; AIDS = acquired immunodeficiency syndrome; and CI = confidence interval.

creases in most types of cancers. The spectrum of HIV-associated cancers may further develop as HIV-infected persons survive longer with highly active antiretroviral therapies. Determining the mechanisms of the specific cancers increased with HIV infection will likely advance the general understanding of cancer etiology.

References

(1) Rabkin CS, Hilgartner MW, Hedberg KW, Aledort LM, Hatzakis A, Eichinger S, et al. Incidence of lymphomas and other cancers in HIV-infected and HIV-uninfected patients with hemophilia. JAMA 1992;267:1090–4.

(2) Vermund SH, Kelley KF, Klein RS, Feingold AR, Schreiber K, Munk G, et al. High risk of human papillomavirus infection and cervical squamous intraepithelial lesions among women with symptomatic human immunodeficiency virus infection. Am J Obstet Gynecol 1991;165:392–400.

(3) Williams AB, Darragh TM, Vranizan K, Ochia C, Moss AR, Palefsky JM. Anal and cervical human papillomavirus infection and risk of anal and cervical epithelial abnormalities in human immunodeficiency virus-infected women. Obstet Gynecol 1994;83:205–11.

(4) Klein RS, Ho GY, Vermund SH, Fleming I, Burk RD. Risk factors for squamous intraepithelial lesions on Pap smear in women at risk for human immunodeficiency virus infection. J Infect Dis 1994;170:1404–9.

(5) Ho GY, Burk RD, Fleming I, Klein RS. Risk of genital human papillomavirus infection in women with human immunodeficiency virus-induced immunosuppression. Int J Cancer 1994;56:788–92.

(6) Rabkin CS, Biggar RJ, Baptiste MS, Abe T, Kohler BA, Nasca PC. Cancer incidence trends in women at high risk of human immunodeficiency virus (HIV) infection. Int J Cancer 1993;55:208–12.

(7) Chiasson MA, Kelley KF, Williams R, Mikl J, Forlenza S, Smith PF. Invasive cervical cancer (ICC) in HIV+ women in New York City (NYC). 3rd Conf Retro and Opportun Infect 130, 1996 [abstract].

(8) Cote TR, O'Brien TR, Ward JW, Wilson SE, Blattner WA. AIDS and cancer registry linkage: measurement and enhancement of registry completeness. The National AIDS/Cancer Match Study Group. Prev Med 1995;24:375–7.

(9) Palefsky JM, Gonzales J, Greenblatt RM, Ahn DK, Hollander H. Anal intraepithelial neoplasia and anal papillomavirus infection among homosexual males with group IV HIV disease. JAMA 1990;263:2911–6.

(10) Kiviat NB, Critchlow CW, Holmes KK, Kuypers J, Sayer J, Dunphy C, et al. Association of anal dysplasia and human papillomavirus with immunosuppression and HIV infection among homosexual men. AIDS 1993;7:43–9.

(11) Rabkin CS, Yellin F. Cancer incidence in a population with a high prevalence of infection with human immunodeficiency virus type 1. J Natl Cancer Inst 1994;86:1711–6.

(12) Hessol NA, Katz MH, Liu JY, Buchbinder SP, Rubino CJ, Holmberg SD. Increased incidence of Hodgkin disease in homosexual men with HIV infection. Ann Intern Med 1992;117:309–11.

(13) Koblin BA, Hessol NA, Zauber AG, Taylor PE, Buchbinder SP, Katz MH, et al. Increased incidence of cancer among homosexual men, New York City and San Francisco, 1978–1990. Am J Epidemiol 1996;144:916–23.

(14) Reynolds P, Saunders LD, Layefsky ME, Lemp GF. The spectrum of acquired immunodeficiency syndrome (AIDS)-associated malignancies in San Francisco, 1980–1987. Am J Epidemiol 1993;137:19–30.

(15) Lyter DW, Kingsley LA, Rinaldo CR, Bryant J. Malignancies in the multicenter AIDS cohort study (MACS), 1984–1994 (Meeting abstract). Proc ASCO 1996;15:A852.

(16) Grulich A, Wan X, Law M, Coates M, Kaldor J. Rates of non-AIDS defining cancers in people with AIDS. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A18 [abstract].

(17) Chadwick EG, Connor EJ, Hanson IC, Joshi VV, Abu-Farsakh H, Yoger R, et al. Tumors of smooth-muscle origin in HIV-infected children. JAMA 1990;263:3182–4.

(18) Steel TR, Pell MF, Turner JJ, Lim GH. Spinal epidural leiomyoma occurring in an HIV-infected man. Case report. J Neurosurg 1993;79:442–5.

(19) Ateenyi-Agaba C. Conjunctival squamous-cell carcinoma associated with HIV infection in Kampala, Uganda. Lancet 1995;345:695–6.

Note

[1]*Editor's note:* SEER is a set of geographically defined, population-based central tumor registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Each registry annually submits its cases to the NCI on a computer tape. These computer tapes are then edited by the NCI and made available for analysis.

Papillomaviruses and Cervical Cancer: Pathogenesis and Vaccine Development

*Douglas R. Lowy, John T. Schiller**

A subset of human papillomaviruses (HPVs) has been implicated as the principal etiologic agents of cervical cancer. Cervical cancers consistently retain and express two of the viral genes, E6 and E7. Although infection with HPV seems to be necessary, other factors, such as cellular immune function, play an important role in determining whether cervical infection will regress, persist, or progress to cancer. The close relationship between viral infection and cancer makes HPV an attractive target for prophylactic and therapeutic vaccines. Candidate vaccines have been shown to have efficacy in animal models, and human clinical trials are planned or in progress. [Monogr Natl Cancer Inst 1998;23:27–30]

In addition to inducing benign papillomas of the skin and mucous membranes, some human papillomaviruses (HPVs) are clearly associated with the development of malignant epithelial tumors *(1–3)*.[1] These cancers include anogenital cancers, especially cancer of the cervix, which is the second most common cancer among women worldwide. A wealth of epidemiologic and molecular biologic data now points to an etiologic link between HPV infection and most cervical cancers. Recognition of the clinical importance of papillomaviruses has stimulated efforts to develop vaccines that may treat or prevent benign and malignant diseases associated with papillomavirus infection.

This article considers the relationship between papillomavirus infection and cervical cancer and describes recent approaches to papillomavirus vaccines.

Subset of Genital/Mucosal HPV Types Found in Cervical Cancer

The HPV replicative cycle is limited to stratified squamous epithelia, with no viremic phase *(4)*. Most of the virus gene expression and replication take place in suprabasal cells that are undergoing terminal differentiation, a strategy that seems designed to evade immune surveillance. More than 70 different HPV types have been identified on the basis of the sequence divergence between their DNA genomes *(2)*. Their genomes, which are a closed, circular, double-stranded DNA approximately 8 kilobases in length, share a similar structural and genetic organization *(5)*. Only a subset of the HPV types appears to regularly infect the genital epithelia. Some of these so-called genital/mucosal types, such as HPV6 and HPV11, are almost never found in cervical cancer and have thus been designated "low-risk" viruses. Others, such as HPV16 and HPV18, are found regularly in cervical cancer and have therefore been designated "high-risk" HPV types *(2,3,6)*.

Many different HPV types have been found in cervical cancer. However, a recent analysis of almost one thousand cervical cancers from different regions of the world *(6)* showed that HPV16 was consistently the most common type, being present in about one half of the cancers from any region. HPV16, HPV18, HPV31, or HPV45 was detected in about 80% of the cancers in every region.

HPV infection of the cervix precedes the onset of cancer by many years. Reliable epidemiologic evidence has shown that HPV infection is by far the most important risk factor (10-fold to 200-fold, compared with controls) for the development of cervical dysplasias, from which almost all cervical cancers arise *(3,7–9)*. The peak in cervical cancer incidence is more than 20 years after the peak in incidence of high-risk genital HPV infection, suggesting that infection per se is insufficient to cause cancer. In most women, cervical infection even with a high-risk HPV is self-limited. Low-risk viruses do not seem to possess the intrinsic capacity to induce cervical cancer.

Inactivation by High-Risk HPVs of Proteins Controlling Cell Proliferation

Considerable progress has been made in identifying potentially important differences between high-risk and low-risk genital/mucosal HPVs *(10)*. Normal human keratinocytes in culture have a finite life span, which is not increased when low-risk HPVs are introduced into them. However, high-risk types, in contrast to low-risk types, can reproducibly induce the immortalization of the keratinocytes. This ability to be passaged indefinitely is a characteristic of partial cellular transformation.

Further analysis of high-risk HPV indicated that keratinocyte immortalization requires two of the viral genes, E6 and E7, which are the same two viral genes that are preferentially retained and expressed in cervical tumors *(11,12)*. Biochemical analysis of the proteins encoded by E6 and E7 has shown that they form complexes with and inactivate specific cellular proteins that probably contribute to the limited life span of cultured keratinocytes and inhibit cell growth *(10,13,14)*. These activities are much lower or are lacking in the E6 and E7 of low-risk viruses. The biological significance of such complexes with high-risk E6 and E7 has been shown most clearly for E6 and the p53 protein and for E7 and the pRB protein, which is encoded by the retinoblastoma tumor susceptibility gene RB. Since p53 and pRB normally control cell proliferation, abrogating these functions places the cells at much greater risk of malignant progression.

**Affiliation of authors:* Laboratory of Cellular Oncology, Division of Basic Sciences, National Cancer Institute, Bethesda, MD.

Correspondence to: Douglas R. Lowy, M.D., National Institutes of Health, Bldg. 36, Rm. 1D32, Bethesda, MD 20892.

See "Note" following "References."

As in the clinical situation, cultured keratinocytes that have been recently immortalized by high-risk HPV are often resistant to differentiation signals but are not fully transformed. Continued passage or the addition of an activated ras oncogene can, in some instances, render these cells tumorigenic for nude mice *(15)*.

Additional Changes Required for Tumorigenic Progression

The long interval between HPV infection and the development of cervical cancer suggests that factors other than viral infection are required for progression to high-grade dysplasia and tumor development. Both virus-specific factors and immune reactivity appear to be important.

While the viral DNA in benign lesions remains extrachromosomal, it is integrated into the host genome in most cervical cancers. During progression, viral expression is usually limited to E6 and E7. Other changes associated with progression likely involve E6- and E7-induced chromosome instability that results from deregulation of cellular growth-control genes and telomerase activity *(16)*.

Persistence of viral infection is associated with progression *(17)*. The likelihood of progression to invasive cancer appears to depend in part on the relative oncogenic potential of the HPV type *(18)*. Some evidence also suggests that genotypie variants within a high-risk type may affect the potential for progression to high-grade dysplasia *(19)*. It remains to be determined whether these apparent differences in likelihood of progression mainly represent biological differences between the viruses or differences in host response.

Cellular immunity to the viral infection seems to represent a critical determinant of whether dysplastic lesions will develop, regress, persist, or progress. HPV16-infected patients with more severe cervical cytology are less likely to show positive cellular immune responses to E6 and E7 antigens than are patients with less severe cytology or with a history of previous HPV16 infection *(20,21)*. There is some evidence that loss of major histocompatibility complex (MHC) class I expression may allow some lesions to evade immune surveillance and progress more rapidly *(22)*. Progression is more common in long-term renal transplant patients on immunosuppressive therapy or in women who are human immunodeficiency virus (HIV) positive, indicating the important role of cellular immune function in host defense *(3)*.

Vaccination Against HPV Infection

The recognition that HPV infection plays the central etiologic role in cervical cancer has fostered efforts to develop vaccines against HPV. Both prophylactic and therapeutic forms of vaccines are under development *(23–28)*. They seek, respectively, to prevent infection or to induce regression of established infection via immune recognition of specific HPV-encoded proteins or peptides. Such vaccines can be delivered either directly as protein, as DNA that encodes and expresses the requisite viral protein(s), or by heterologous viral vectors *(29)*.

General Considerations

A major barrier to developing a practical vaccine is that most critical HPV-immune determinants are likely to be type specific because the proteins of different HPV types are quite divergent at the amino acid level. This limitation implies that protection induced by protein from a given HPV type is likely to be type specific. It will therefore be necessary either to have a specific vaccine for each HPV type or to incorporate viral protein from an appropriate spectrum of HPV types in a polyvalent vaccine.

Further difficulties are that HPV does not cause disease in animals and is not infectious for them. This means that animal studies must be carried out with animal papillomaviruses, with grafted human material in immunologically suppressed hosts, or with model systems that incorporate specific HPV genes or proteins. It is not always clear whether experimental results obtained with these animal models will apply directly to clinical HPV infection in humans.

Prophylactic Vaccines

Encouraging results have come from animal studies of vaccines to prevent papillomavirus infection. Consistent protection (90%–100%) has been obtained by immunizing animals with virus-like particles (VLPs) composed of the major structural viral protein L1 *(30–32)*.

The papillomavirus capsid is primarily composed of 360 molecules of L1 protein, and L1, when expressed in the absence of other papillomavirus genes, can self-assemble into VLPs that are morphologically and immunologically similar to infeetious papillomavirus *(33,34)*. Since the VLPs are produced by genetically engineered cells that do not contain the nonstructural viral genes, such as E6 or E7, the VLPs do not contain the papillomavirus DNA genome, are not infectious, and cannot cause neoplastic changes in cells.

Immunization with VLPs *(30)* or with the L1 gene *(35)* has produced substantial protection in rabbits against experimental challenge with the Shope cottontail rabbit papillomavirus (CRPV), which induces cutaneous papillomas that can progress to malignant squamous cell carcinomas. VLP immunization can also prevent experimental oral mucosal infection in dogs by canine oral papillomavirus *(31)* and in cows by bovine papillomavirus type 4 (BPV4) *(32)*.

In these studies, the protection, which can be passively transferred by antibodies from immune animals to nonimmune animals, is mediated by antibodies directed against conformational epitopes that are present on the VLP as well as on infectious papillomaviruses. Since the conformational epitopes are type specific *(36–38)*, protection is type specific *(30)*. The efficacy of these protocols seems to be limited to prevention; BPV4 VLP immunization did not induce regression of established BPV4 papillomas *(32)*.

These promising animal studies are leading to the testing of a candidate prophylactic HPV vaccine in humans. In addition to deciding which HPV types to include in such a vaccine, it will be desirable to optimize the adjuvant, dosage, and route of administration. Efforts to improve the mucosal immunity induced by VLPs may increase their efficacy against mucosal genital infection.

However, there are reasons to believe that preventing genital mucosal HPV infection may not require the induction of mucosal immunoglobulin (Ig) A. In addition to the animal studies that show protection against experimental oral infection, transudation of IgG into vaginal secretions can be induced by systemic

immunization with purified protein *(39)*. Furthermore, it is believed that initiation of genital HPV infection occurs only when there is sufficient trauma to allow the virions to come in contact with the proliferating epithelial cells, which are located in the basal layer of the epithelium. This hypothesis is supported by the tentative identification of α_6 integrin, which is not expressed in suprabasal cells, as a candidate receptor for papillomaviruses *(40)*. It is reasonable to expect that trauma which was sufficient to abrade the epithelium would usually be associated with exudation of systemic IgG into the abraded area, where it might neutralize the HPV virions.

Therapeutic Vaccines

Therapeutic vaccines might be used in various settings, including the treatment of invasive cancers, as adjunct therapy to prevent recurrence or metastasis, against dysplasias, or in benign disease. A major theoretical obstacle to developing such vaccines is that the immunologic determinants for viral persistence or regression remain poorly defined *(41,42)*, although it is clear that patients with impaired cellular immunity are at increased risk of persistent HPV infection and carcinogenic progression. Most efforts have been directed toward using the E6 and E7 proteins, or peptides derived from them, largely because these are the viral proteins that are retained and expressed in cervical tumors *(43)*. However, vaccines directed against benign lesions would not need to be limited to these two proteins. For example, E1 and E2, which are required for the viral DNA to be maintained as an extrachromosomal element that is unintegrated in the host DNA, represent potentially interesting targets for benign lesions.

Since these papillomavirus proteins are not expressed on the cell surface, there is little potential for antibody-dependent cytotoxicity to mediate regression. Instead, potentially effective cytotoxic responses will probably require a vaccine that induces the presentation of small virally encoded peptides to antigen-presenting cells. In cells that possess class I molecules, the normal process of partial intracellular degradation of cytoplasmic or nuclear viral proteins can, following the binding of small viral peptides to the class I molecules, lead to the induction of antigen-specific reactivity of CD8-positive cytotoxic T lymphocytes (CTLs).

Introduction of the relevant viral gene or protein into target cells is normally required to develop virus-specific CTLs. Papillomavirus genes can be introduced into cells as naked DNA, which is a relatively inefficient process, or as part of a viral vector such as vaccinia virus, which is usually more efficient. Rodents immunized with HPV16 E6 or HPV16 E7, delivered either in vaccinia virus vectors containing one of the two viral genes or in killed tumor cells that expressed one of the genes and the gene for immune co-stimulatory protein B7, were protected against subsequent challenge with tumor cells expressing the corresponding viral protein *(44–46)*. A recombinant vaccinia human safety trial of an HPV16 and HPV18 E6 and E7 virus has been completed *(47)*, and others are under way or planned *(28)*.

A CTL response can also be induced by the viral protein itself if it is taken up by cells in a manner that leads to its partial degradation and presentation with class I antigen. Although soluble protein by itself does not usually generate such a response, injection of rabbits with bacterially derived CRPV E1 or E2 protein was shown to increase the rejection rate of CRPV-induced papillomas *(48)*. The precise immunologic mechanism underlying this effect remains to be established. However, class I-dependent CD8 CTLs can be reproducibly induced if viral protein is presented in particulate form *(49)*, as part of a VLP, in combination with some adjuvants, or encapsidated into liposomes.

Another alternative is to immunize with small viral peptides, which, following their binding to empty class I molecules on cell surfaces, can induce cytotoxic CD8 T cells *(50)*. A difficulty associated with this approach is that the immunogenicity of a given peptide is genetically determined by the class I alleles present in a given individual, which means that only some individuals will respond to even an "immunogenic" peptide. Also, small peptides are often quite unstable *in vivo (51)*.

In some tumor models, immunologic rejection is mediated by antigen-specific CD4 T cells rather than by CD8 T cells *(52,53)*. While the class I CD8 pathway is characteristic of cytoplasmic and nuclear proteins, CD4 T cells can be activated by the processing of endogenously expressed membrane-associated proteins for processing and presentation by MHC class II molecules to CD4 cells. To induce E7-specific CD4 cells, a study *(54)* used genetic engineering to target the ordinarily nuclear HPV16 E7 protein to the lysosomal compartment by adding a lysosomal membrane targeting signal to E7. Such a lysosomally targeted E7 protein, which was introduced into mice via a vaccinia virus vector carrying the engineered E7 gene, was highly effective in protecting mice against tumors derived from a mouse epithelial cell line transformed by HPV16 E6 and E7 genes and a ras oncogene *(55)*. In this tumor model, protection by the lysosomal E7 protein was found to depend on both CD4 and CD8 cells. This observation may explain why authentic E7, which presumably induced only active CD8 cells, was much less effective against the tumors than was the lysosomally targeted protein.

Thus, various approaches are being taken to develop HPV vaccines. Information from human clinical trials with candidate vaccines may be expected in the next few years.

References

(1) Lowy DR, Kirnbauer R, Schiller JT. Genital human papillomavirus infection. Proc Natl Acad Sci U S A 1994;91:2436–40.

(2) zur Hausen H. Molecular pathogenesis of cancer of the cervix and its causation by specific human papillomavirus types. Curr Top Microbiol Immunol 1994;186:131–56.

(3) Human papillomaviruses. IARC Monogr Eval Carcinog Risks Hum, vol 64. 1995.

(4) Taichman LB, LaPorta RF. The expression of papillomaviruses in human epithelial cells. In: Salzman NP, Howley PM, editors. The papovaviridae: vol 2. The papillomaviruses. New York: Plenum Press, 1987:109–39.

(5) Turek L. The structure, function and regulation of papillomaviral genes in infection and cervical cancer. Adv Virus Res 1994;44:305–56.

(6) Bosch FX, Manos MM, Munoz N, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide prospective. J Natl Cancer Inst 1995; 87:796–802.

(7) Schiffman M, Bauer HM, Hoover RN, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J Natl Cancer Inst 1993;85:958–64.

(8) Schiffman MH. Epidemiology of cervical human papillomaviruses. In: zur Hausen H, editor. Human pathogenic papillomaviruses. Heidelberg: Springer-Verlag, 1994:55–81.

(9) Munoz N, Bosch FX, de Sanjose S, et al. The role of HPV in the etiology of cervical cancer. Mutat Res 1994;305:293–301.

(10) Werness BA, Munger K, Howley PM. Role of the human papillomavirus oncoproteins in transformation and carcinogenic progression. In: DeVita

VT Jr, Hellman S, Rosenberg SA, editors. Important advances in oncology 1991. Philadelphia: Lippincott, 1991:3–18.
(11) Vousden K. Interactions of human papillomavirus transforming proteins with the products of tumor suppressor genes. FASEB J 1993;7:872–9.
(12) Mansur CP, Androphy AJ. Cellular transformation by papillomavirus oncoproteins. Biochim Biophys Acta 1993;1155:323–45.
(13) Scheffner M, Huibregtse JM, Vierstra RD, et al. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell 1993;75:495–505.
(14) Chen JJ, Reid CE, Band V, et al. Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. Science 1995;269: 529–31.
(15) DiPaolo JA, Woodworth CD, Popescu NC, et al. Induction of human cervical squamous cell carcinoma by sequential transfection with human papillomavirus 16 DNA and viral Harvey ras. Oncogene 1989;4:395–9.
(16) Klingelhutz AJ, Foster SA, McDougall JK. Telomerase activation by the E6 gene product of human papillomavirus type 16. Nature 1996;380:79–82.
(17) Gloria YF, Ho F, Burk RD, et al. Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. J Natl Cancer Inst 1995;87:1365–71.
(18) Lorincz AT, Reid R, Jenson AB, et al. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. Obstet Gynecol 1992;79:328–37.
(19) Xi LF, Koutsky LA, Galloway DA, et al. Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. J Natl Cancer Inst 1997;89:796–802.
(20) Tsukui T, Hildesheim A, Schiffman M, et al. Interleukin 2 production *in vitro* by peripheral lymphocytes in response to human papillomavirus-derived peptides: correlation with cervical pathology. Cancer Res 1996;56: 3967–74.
(21) Nakagawa M, Stites D, Farhat S, et al. Cytotoxic T lymphocyte responses to E6 and E7 proteins of human papillomavirus type 16: relationship to cervical intraepithelial neoplasia. J Infect Dis 1997;175:927–31.
(22) Connor ME, Stern PL. Loss of MHC class-1 expression in cervical carcinomas. Int J Cancer 1990;46:1029–34.
(23) Galloway DA. Human papillomavirus vaccines: a warty problem. Infect Agents Dis 1994;3:187–93.
(24) Hines JF, Ghim S, Schlegel R, et al. Prospects for a vaccine against human papillomavirus. Obstet Gynecol 1995;86:860–6.
(25) Munoz N, Crawford L, Coursaget P. HPV vaccines and their potential use in the prevention and treatment of cervical neoplasia. Papillomavirus Rep 1995;6:54–5.
(26) Schiller JT, Okun M. Papillomavirus vaccines: current status and future prospects. Adv Dermatol 1996;11:355–80.
(27) Frazer I. Strategies for immunoprophylaxis and immunotherapy of papillomaviruses. Clin Dermatol 1997;15:285–97.
(28) Hanissian J. Emerging HPV vaccines. Infect Med 1997;14:266,273–275, 330.
(29) Rabinovich NR, McInnes P, Klein DL, et al. Vaccine technologies: view to the future. Science 1994;265:1401–4.
(30) Breitburd F, Kirnbauer R, Hubbert NL, et al. Immunization with virus-like particles from cottontail rabbit papillomavirus (CRPV) can protect against experimental CRPV infection. J Virol 1995;69:3959–63.
(31) Suzich JA, Ghim S, Palmer-Hill FJ, et al. Systemic immunization with papillomavirus L1 protein completely prevents the development of viral mucosal papillomas. Proc Natl Acad Sci U S A 1995;92:11553–7.
(32) Kirnbauer R, Chandrachud L, O'Neil B, et al. Virus-like particles of bovine papillomavirus type 4 in prophylactic and therapeutic immunization. Virology 1996;219:37–44.
(33) Kirnbauer R, Booy F, Cheng N, et al. Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. Proc Natl Acad Sci U S A 1992;89:12180–4.
(34) Hagensee ME, Yaegashi N, Galloway DA. Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. J Virol 1993;67:315–22.
(35) Donnelly JJ, Martinez D, Jansen KU, et al. Protection against papillomavirus with a polynucleotide vaccine. J Infect Dis 1996;173:314–20.
(36) Christensen ND, Kirnbauer R, Schiller JT, et al. Human papillomavirus types 6 and 11 have antigenically distinct strongly immunogenic conformationally dependent neutralizing epitopes. Virology 1994;205: 329–35.
(37) Roden RB, Hubbert NL, Kirnbauer R, et al. Assessment of the serological relatedness of genital human papillomaviruses by hemagglutination inhibition. J Virol 1996;70:3298–301.
(38) Roden RB, Greenstone HL, Kirnbauer R, et al. *In vitro* generation and type-specific neutralization of a human papillomavirus type 16 virion pseudotype. J Virol 1996;70:5875–83.
(39) Bouvet JP, Belec L, Pires R, et al. Immunoglobulin G antibodies in human vaginal secretions after parenteral vaccination. Infect Immun 1994;62: 3957–61.
(40) Evander M, Frazer IH, Payne E, et al. Identification of the α_6 integrin as a candidate receptor for papillomaviruses. J Virol 1997;71:2449–56.
(41) Tindle RW, Frazer IH. Immune response to human papillomaviruses and the prospects for human papillomavirus-specific immunization. Curr Top Microbiol Immunol 1994;186:217–53.
(42) Dillner J. Serology of human papillomavirus. Cancer J 1995;8:264–9.
(43) Steller MA, Schiller JT. Human papillomavirus immunology and vaccine prospects. Monogr Natl Cancer Inst 1996;21:145–8.
(44) Meneguzzi G, Cerni C, Kieny MP, et al. Immunization against human papillomavirus type 16 tumor cells with recombinant vaccinia virus expressing E6 and E7. Virology 1991;181:62–9.
(45) Chen L, Thomas EK, Hu SL, et al. Human papillomavirus type 16 nucleoprotein E7 is a tumor rejection antigen. Proc Natl Acad Sci U S A 1991; 88:110–4.
(46) Chen L, Mizuno MT, Singhal MC, et al. Induction of cytotoxic T lymphocytes specific for a syngeneic tumor expressing the E6 oncoprotein of human papillomavirus type 16. J Immunol 1992;148:2617–21.
(47) Borysiewicz LK, Fiander A, Nimako M, et al. A recombinant vaccinia virus encoding human papillomavirus types 16 and 18, E6 and E7 proteins as immunotherapy for cervical cancer. Lancet 1996;1:1523–7.
(48) Selvakumar R, Borenstein LA, Lin YL, et al. Immunization with nonstructural proteins E1 and E2 of cottontail rabbit papillomavirus stimulates regression of virus-induced papillomas. J Virol 1995;69:602–5.
(49) Tindle RW, Herd K, Londono P. Chimeric hepatitis B core antigen particles containing B- and Th- epitopes of human papillomavirus type 16 E7 proteins induce specific antibody and T-helper responses in immunized mice. Virology 1994;200:547–57.
(50) Feltkamp MC, Smits HL, Vierboom MP, et al. Vaccination with cytotoxic T lymphocyte epitope-containing peptide protects against a tumor induced by human papillomavirus type 16-transformed cells. Eur J Immunol 1993; 23:2242–9.
(51) Falo LDJ, Colarusso LJ, Benacerraf B, et al. Serum proteases alter the antigenicity of peptides presented by class I major histocompatibility complex molecules. Proc Natl Acad Sci U S A 1992;89:8347–50.
(52) Golumbek PT, Lazenby AJ, Levitsky HI, et al. Treatment of established cancer by tumor cells engineered to secrete interleukin-4. Science 1991; 254:713–6.
(53) Topalian SL, Rivoltini L, Mancini M, et al. Human CD4+ T lymphocytes specifically recognize a shared melanoma-associated antigen encoded by the tyrosine gene. Proc Natl Acad Sci U S A 1994;92:9481–5.
(54) Wu TC, Guarnieri FG, Staveley O, et al. Engineering an intracellular pathway for major histocompatibility complex class II presentation of antigens. Proc Natl Acad Sci U S A 1995;92:11671–5.
(55) Lin KY, Guarnieri FG, Staveley-O'Carroll KF, et al. Treatment of established tumors with a novel vaccine that enhances major histocompatibility class II presentation of tumor antigen. Cancer Res 1996;56:21–6.

Note

[1]Portions of this article are adapted from Lowy DR, Schiller JT. Oncogenesis and vaccine prospects for the papillomaviruses. Curr Opin Dermatol 1997;4: 256–61.

Cancers in Human Immunodeficiency Virus-Infected Children

*Brigitta U. Mueller**

Although the exact incidence of cancers in human immunodeficiency virus (HIV)-infected children is not clear, an excess of non-Hodgkin's lymphomas and soft tissue tumors as well as a multitude of otherwise rare tumors in childhood, such as cervical, thyroid, or pulmonary carcinoma, has been reported. In contrast to the findings in HIV-infected adults, Kaposi's sarcoma is rare in children in industrialized countries but not in children living in the sub-Saharan area. Treatment of the neoplastic disease is often complicated by multiple HIV-associated organ dysfunctions as well as drug interactions and infectious complications secondary to severe immunosuppression. Nonetheless, preliminary results with dose-intensive, but brief, chemotherapeutic regimens have been encouraging, and HIV-infected children who develop cancer are likely to benefit from aggressive treatment combined with adequate supportive care. Furthermore, insights gained from the study and treatment of this very challenging group of patients may benefit other immunocompromised hosts as well as increase our understanding of oncogenesis in general. [Monogr Natl Cancer Inst 1998;23: 31–35]

Background and Epidemiology

Worldwide, the Joint United Nations Programme on AIDS [acquired immunodeficiency disease syndrome] (UNAIDS) estimates that 400 000 new human immunodeficiency virus (HIV) infections occurred in children in 1996 and that currently 830 000 children are living with symptoms of HIV/AIDS (UNAIDS Web site). Of the 2.6 million children who were infected since the epidemic began, an estimated 1.4 million have already died. The number of children with HIV infection who develop a cancer is poorly defined. Among the 7629 children diagnosed with AIDS by the end of December 1996, 156 (2.04%) had a cancer as their AIDS-defining diagnosis *(1)*. However, the AIDS definition revised by the Centers for Disease Control and Prevention (CDC) in 1994 lists only a limited number of neoplasms: Kaposi's sarcoma (KS) lesions, primary non-Hodgkin's lymphomas (NHLs) of the central nervous system (CNS), small non-cleaved cell lymphomas, and immunoblastic large-cell NHLs of B-cell or unknown phenotype *(2)*. Leiomyosarcomas are listed under category B of the CDC classification, representing a "moderate" clinical symptom. Furthermore, children who meet the diagnosis of AIDS with another symptom before they develop a tumor will not be registered again. It is therefore clear that the CDC registry will provide only limited information about the true incidence of neoplasms in HIV-infected children.

Among the 156 tumors captured in the CDC AIDS definition in children, there were 50 Burkitt's lymphomas, 48 immunoblastic lymphomas, 30 NHLs of the CNS, and 28 KS lesions. The number of KS cases is probably an overestimation because of initial problems with the pathologic diagnosis (Lindegren ML [CDC]: personal communication). If one takes into account only these numbers, a conservative estimate is that children with HIV infection appear to have at least a 100-fold higher incidence of cancers. In 1996 alone, 10 of 678 children who were newly diagnosed with AIDS received this diagnosis because of an NHL or a KS. This represents an incidence of neoplasms in children with AIDS of 1.47%, compared with healthy, non-HIV-infected children, who develop cancer in approximately 130 cases per million children (0.013%) per year *(1,3)*.

Several efforts have been made to better define the number and range of neoplastic manifestations in HIV-infected children. In a combined, retrospective survey performed by the HIV and AIDS Malignancy Branch of the National Cancer Institute (NCI), Bethesda, MD, and the Children's Cancer Group, 59 tumors were identified in 57 children (Granovsky M [NCI]: personal communication). It is interesting that 42% of these tumors would not have been captured by the CDC classification. These data are very comparable to a prospective survey performed by the Pediatric Oncology Group *(4)*. The tumors observed are often unusual and extremely rare in non-HIV-infected children; they include carcinoma of the thyroid, oat cell tumor of the tracheo–esophageal region, carcinomas of the vagina and cervix, hepatoblastoma, and soft tissue tumors (e.g., leiomyosarcomas), among others *(5–8)*. Several reports from Africa have noted an increased incidence of retinoblastomas, nasopharyngeal carcinomas, and rhabdomyosarcomas, tumors not commonly associated with immune deficiency states *(9,10)*.

There are clear differences between HIV-infected children and HIV-infected adults. For example, KS is a rare tumor in children in industrialized countries (United States and Europe), perhaps with the exception of Haiti and Romania *(11)*. It is interesting that KS is the AIDS-defining illness in only 3% of adolescents between 13 and 19 years of age, but this incidence increases to 9% in young adults between 20 and 24 years of age and to 13% in adults older than 25 years *(1,12)*. KS is more common in homosexual HIV-infected men than in heterosexual HIV-infected men or women. However, there is an increasing incidence of KS among children living in Africa *(9,10,13–15)*. A

**Correspondence to:* Brigitta U. Mueller, M.D., Assistant Professor of Pediatrics, Harvard Medical School, Children's Hospital, HU-215, 300 Longwood Ave., Boston, MA 02115. E-mail: mueller_b@a1.tch.harvard.edu

© Oxford University Press

further difference between adults and children with HIV infection is the increased incidence in children of smooth muscle tumors, leiomyomas, and leiomyosarcomas *(8,16–20)* as well as the high prevalence in children of lymphoproliferative disorders (Table 1) *(21–24)*. It is currently not clear whether there is an increased tendency of these lymphoproliferative disorders to evolve into a "true," monoclonal cancer with a rapid and invasive growth pattern.

Non-Hodgkin's Lymphoma

As indicated above, the CDC classification includes small, non-cleaved cell lymphomas (mainly B-cell phenotype), large-cell lymphomas (immunoblastic and anaplastic tumors), and primary CNS lymphomas.

In non-HIV-infected children, there are two forms of small, non-cleaved cell NHLs (Burkitt's and Burkitt's-like lymphomas). The endemic form, more than 95% of which is associated with Epstein-Barr virus infection, presents typically with jaw tumors and has an especially high prevalence in equatorial Africa. The sporadic form is observed in the United States, Europe, and Asia and is 20- to 100-fold less common than the endemic form *(25)*. Burkitt's lymphoma constitutes about 50% of the childhood lymphomas. It has an annual incidence in children of 76.2 per million in Nigeria compared with 0.3 (U.S. blacks) and 2.0 (U.S. whites) per million in the United States. However, either form is rare compared with an extrapolated annual incidence of more than 4400 cases per million children with AIDS (based on the CDC statistics of three of 678 children diagnosed with AIDS because of a Burkitt's NHL in 1996) *(1,12)*.

In non-HIV-infected children, only about 20% of all NHLs are of the large-cell type, and most of them are of the B-cell phenotype *(26)*. The incidence of large-cell tumors appears to be higher in HIV-infected children, and anaplastic T-cell NHL can occur. Among the 13 HIV-infected children with NHL followed at the NCI, five (38%) had large-cell tumors. Two were immunoblastic B-cell tumors, two were anaplastic T-cell tumors, and one was an anaplastic tumor of non-B, non-T phenotype. Lymphoblastic NHLs, which represent about one third of all NHLs in children in the United States, are predominantly of T-cell origin and do not appear to be more common in the HIV-infected population. They are not included in the CDC AIDS definition.

Table 1. Lymphoproliferative disorders in human immunodeficiency virus-infected children*

Disorder	Comments
Polyclonal hypergammaglobulinemia	Common
Reactive lymphadenopathy	Common
Lymphocytic interstitial pneumonitis	Common, can lead to oxygen and steroid dependency
Mucosa-associated lymphoid tumors	Lungs, gastric mucosa, and parotid or lacrimal glands
Cystic mediastinal tumors	Polyclonal, often indolent
Diffuse infiltrative lymphocyte syndrome	Parotid or lacrimal glands
Lymphomatoid papulosis	Monoclonal, indolent, T-cell process with waxing and waning course
Castleman's disease	Rare in children

*The distinction among the different manifestations is not always well defined. Most of these processes are polyclonal or oligoclonal, but they might evolve into a monoclonal process.

Although secondary involvement of the CNS is observed, especially in children with bone marrow involvement, it is extremely rare to see a primary CNS lymphoma in the non-HIV-infected child *(26)*. However, it is the AIDS-defining diagnosis in 0.4% of children, constituting 23% of all the NHLs registered for AIDS diagnosis. However, in 1996, only one (10%) of 10 children had a CNS NHL as an AIDS-defining tumor, compared with 230 (22%) of 1030 adults *(1)*. This finding possibly reflects a change in the disease course in children, since CNS NHLs are most prevalent in patients with very low CD4 counts and have often been diagnosed at autopsy. As in adults, the differential diagnosis, especially in the adolescent, has to include cerebral toxoplasmosis. However, we also have observed at least one case of an intracranial lymphoproliferative lesion similar to what has been described in the post-transplant situation. Furthermore, since HIV-infected children are now surviving longer, it has to be kept in mind that brain tumors are the most common solid tumors in otherwise healthy children *(3)*. An empirical treatment for toxoplasmosis may not be warranted in the younger child, since cerebral toxoplasmosis with mass lesion(s) is extremely rare in preadolescent children, and an early biopsy should be considered.

Unusual, extranodal sites of presentation of NHLs are relatively common in the HIV-infected child (Table 2) *(27–31)*. Intriguing is the high incidence of mucosa-associated lymphoid tumors, observed in pulmonary or gastric mucosa as well as in the parotid, salivary, or lacrimal glands *(29,32,33)*. These oligoclonal or monoclonal processes are usually curable by surgical excision alone and may represent an interface between lymphoproliferative disorder and cancer. Lymphocytic interstitial pneumonitis is a common manifestation of pediatric HIV disease, and it is possible that pulmonary mucosa-associated lymphoid tumors and lymphocytic interstitial pneumonitis are related *(34–37)*.

Kaposi's Sarcoma

KS is rarely diagnosed in HIV-infected children from the United States, with the exception of vertically infected children from Haiti or older adolescents *(38–42)*. However, in African countries such as Zambia or Uganda, KS now constitutes almost 20% of all childhood cancers, compared with 6% before 1986, and the male-to-female ratio has changed from 3:1 before the AIDS epidemic to about 1.7:1 at the present time *(9,10,15)*. It is interesting that this more than 40-fold increase in the occurrence

Table 2. Unusual sites of non-Hodgkin's lymphomas in the human immunodeficiency virus-infected child

Site	Comment
Lungs	Presenting as cavitary lesion or as mucosa-associated lymphoid tumor Reminiscent of lymphomatoid papulosis
Heart	Rare
Liver	Differential diagnosis with lymphoproliferative disorder
Gastrointestinal tract	In gastric mucosa presenting as mucosa-associated lymphoid tumor
Bone	Bony lesions without overt bone marrow involvement
Central nervous system	As single or multiple lesions

of childhood KS, although mainly occurring in HIV-infected children, has also been observed in noninfected pediatric populations. The lymphadenopathic form is most prevalent, and common sites include head and neck (82%), extremities (7%), inguinal (5%), and abdomen (4%). In adults, human herpesvirus type 8 (HHV-8) has been shown to be present in more than 90% of all KS lesions, as well as in AIDS-related body cavity NHLs, lymphomatous effusions without tumor mass, and Castleman's disease *(43–46)*. In the general population of the United States, about 25% of adults have HHV-8-related antibodies, compared with 18% of adolescents and less than 4% of children under 13 years of age *(47–51)*. However, in African countries such as Nigeria or Zaire, the seroprevalence in the population is much higher (56% and 82%, respectively). In a study from Uganda, nine of 18 children with KS were also positive for HHV-8, compared with seven of 22 children with other tumors *(9,15)*.

Smooth Muscle Tumors

Leiomyomas and leiomyosarcomas are extremely rare in non-HIV-infected children, but at least 18 cases have now been observed in HIV-infected children. While leiomyomas can be incidental findings and can remain clinically indolent, leiomyosarcomas often present as widespread disease. As described for lymphomas, smooth muscle tumors can also occur in unusual locations (e.g., in the lungs, adrenal glands, or skin) *(8,16–20,52,53)*. Leiomyosarcomas from children with HIV infection as well as from patients after receiving a transplant but not from other patients have been shown to be positive for Epstein-Barr virus by *in situ* hybridization *(20,54)*. Quantitative polymerase chain reaction showed relatively high levels of Epstein-Barr virus in tumor tissue (as many as 4.3 genome copies per cell), and discrete episomal Epstein-Barr virus clones were found in lesions taken from different sites in the same patient, indicating simultaneous emergence of separate clones *(20)*. It is currently unclear whether HIV infection and the associated disturbance in immune surveillance lead to an increased expression of the Epstein-Barr virus receptor or whether Epstein-Barr virus infection of the smooth muscle cells results in increased expression.

Miscellaneous Tumors

Anal *(55,56)* and cervical *(57,58)* carcinomas might be seen in sexually active adolescents, since infection with human papillomavirus, which is known to be associated with the oecurrence of these tumors, is frequently acquired by adolescents *(59)*. Hodgkin's disease is more common in adolescents than in younger children and, as in adults, it is possible that a slight increase in the number of this relatively common tumor will be observed in the future *(60)*. Hepatic tumors might become more prevalent, since many HIV-infeeted children, especially those who are hemophiliacs, are also infected with hepatitis B or C *(61–63)*.

HIV-infected children (like HIV-infected adults) are surviving longer and are in better health, thanks to improved and expanded therapeutic options. It is therefore possible that more children in the future will develop the "normal" pediatric cancers, especially leukemia and brain tumors. While these tumors may not be more common in the HIV-infected child, their diagnosis and therapy will certainly pose unique problems.

Children are being exposed to multiple drugs for the treatment of HIV disease and its complications not only during childhood but also prenatally. To decrease the transmission rate of HIV disease, it is currently recommended that all pregnant HIV-positive women receive zidovudine during pregnancy and birth and that their infants be treated with zidovudine for the first 6 weeks of life *(64,65)*. Recently, studies of the offsprings of mice that had been treated with zidovudine during the last trimester *(66)* revealed an increased risk of developing liver and lung tumors as well as tumors of the reproductive organs. Olivero et al. have demonstrated that zidovudine is incorporated into the DNA of newborn mice and monkeys, as well as into the nuclear DNA of cord blood samples drawn from children whose mothers had been treated with the drug. In January 1997, the National Institutes of Health (Bethesda, MD) convened a panel to review these studies and, although acknowledging the validity of the findings, also recognized that the benefit of preventing transmission of HIV disease in the vast majority of children currently outweighs the potential concerns of carcinogenicity. However, the panel also strongly emphasized the need for careful, long-term follow-up of *all* children exposed *in utero* to antiretroviral therapy, including those children who are not HIV infected.

Present and Future Challenges

There are several dilemmas that the physician who is treating HIV-infected children is facing. First, as mentioned, the scope of the problem, including the exact incidence and spectrum of tumors, is not well defined. It may be necessary in the future to adapt the CDC definition, and it clearly remains important to use the AIDS malignancy registries as well as to publicize unusual cases in order to capture the (potentially very broad and unexpected) spectrum of pediatric AIDS-related tumors.

Second, the treating physician is faced with diagnostic dilemmas. Atypical symptoms and courses can occur, unusual cancers may be encountered, and the differential diagnosis is often confusing because of similar presentations of opportunistic infections (e.g., CNS lesion and toxoplasmosis or fever and night sweats in disseminated *Mycobacterium avium* complex infection). Furthermore, even if a child is successfully treated for one tumor, he or she remains at risk for another cancer as long as the immunosuppression continues.

Thirdly, there are therapeutic dilemmas. Children are often followed by an infectious disease specialist who has no oncology experience or at a cancer center that has no HIV experience, and an inter-disciplinary approach appears certainly more advantageous. Furthermore, there are currently no standardized treatment guidelines or protocols available for HIV-infected (or otherwise immune compromised) children who develop a cancer. Many of these children have unique problems, such as preexisting organ damage, especially diminished bone marrow reserves, and concurrent opportunistic infections, and they are also receiving multiple drugs (antiretroviral and other) that might lead to potentially dangerous drug interactions. However, there is increasing evidence that children with HIV infection, despite the multiple problems involved, are able to tolerate even intensive chemotherapy (at least for limited periods of time) and can be cured of their cancer without acceleration of their underlying disease *(31,67–69)*. One such example is the dose-intensive protocol currently used by the NCI for the treatment of NHL in

immunocompromised children *(31)*. A combination of cytoxan (day 1) and methotrexate (day 10) is given for three cycles as soon as the absolute neutrophil count reaches 500 cells/mm^3. Seven of nine patients achieved a complete remission, and no patient died during or as a result of chemotherapy.

In summary, there are many unsolved problems to be addressed regarding the epidemiology, etiology, diagnosis, and therapy of cancers occurring in HIV-infected children. Knowledge gained in studying these questions not only will benefit the affected children but also will further our understanding of oncogenesis and of the interactions between viral infections and the immune system, and shorter but still effective therapies might benefit the nonimmune compromised cancer patient as well.

References

(1) Centers for Disease Control and Prevention. U.S. HIV and AIDS cases reported through December 1996. HIV/AIDS Surveillance Report 1996;8: 1–39.

(2) Centers for Disease Control and Prevention. Revised classification system for human immunodeficiency virus infection in children less than 13 years of age. MMWR Morb Mortal Wkly Rep 1994;43:1–12.

(3) Robison LL. General principles of the epidemiology of childhood cancer. In: Pizzo PA, Poplack DG, editors. Principles and practice of pediatric oncology. Philadelphia: Lippincott-Raven Publishers, 1997:1–10.

(4) Pollock BH, Jenson HB, McClain KL, et al. Risk factors for HIV-related malignancies in children. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A17.

(5) Diamond FB, Price LJ, Nelson RP Jr. Papillary carcinoma of the thyroid in a seven-year-old HIV-positive child. Pediatr AIDS and HIV Infect: From Fetus to Adolescent 1994;5:232–5.

(6) Ninane J, Moulin D, Latinne D, et al. AIDS in two African children—one with fibrosarcoma of the liver. Eur J Pediatr 1985;144:385–90.

(7) Scully RE, Mark RE, McNeely BU. Case records of the Massachusetts General Hospital. N Engl J Med 1986;314:629–40.

(8) Chadwick EG, Connor EJ, Guerra Hanson IC, et al. Tumors of smooth muscle origin in HIV-infected children. JAMA 1990;263:3182–4.

(9) Ziegler JL, Katongole-Mbidde E. Kaposi's sarcoma in childhood: an analysis of 100 cases from Uganda and relationship to HIV infection. Int J Cancer 1996;65:200–3.

(10) Athale UH, Patil PS, Chintu C, et al. Influence of HIV epidemic on the incidence of Kaposi's sarcoma in Zambian children. J Acquir Immune Defic Syndr Hum Retrovirol 1995;8:96–100.

(11) Orlow SJ, Cooper D, Petrea S, et al. AIDS-associated Kaposi's sarcoma in Romanian children. J Am Acad Dermatol 1993;28:449–53.

(12) Centers for Disease Control and Prevention. AIDS among children—United States, 1996. MMWR Morb Mortal Wkly Rep 1996;45:1005–10.

(13) Davachi F, Ngoie K, Ndoko K, et al. Pediatric AIDS and associated cancers in Kinshasa, Zaire. Antibiot Chemother 1991;43:140–9.

(14) Chintu C, Athale UH, Patil PS. Childhood cancers in Zambia before and after the HIV epidemic. Arch Dis Child 1995;73:100–5.

(15) Mbulaiteye SM, Ziegler JL, Katongole-Mbidde E, et al. Risk factors for childhood Kaposi's sarcoma in Uganda: a case control study. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A18.

(16) McLoughlin LC, Nord KS, Joshi VV, et al. Disseminated leiomyosarcoma in a child with acquired immune deficiency syndrome. Cancer 1991;67: 2618–21.

(17) Sabatino D, Martinez S, Young R, et al. Simultaneous pulmonary leiomyosarcoma and leiomyoma in pediatric HIV infection. Pediatr Hematol Oncol 1991;8:355–9.

(18) Orlow SJ, Kamino H, Lawrence RL. Multiple subcutaneous leiomyosarcomas in an adolescent with AIDS. Am J Pediatr Hematol Oncol 1992;14: 265–8.

(19) Ross JS, Del Rosario A, Bui HX, et al. Primary hepatic leiomyosarcoma in a child with the acquired immunodeficiency syndrome. Hum Pathol 1992; 23:69–72.

(20) McClain KL, Leach CT, Jenson HB, et al. Association of Epstein-Barr virus with leiomyosarcomas in young people with AIDS. N Engl J Med 1995;332:12–8.

(21) Joshi VV. Systemic lymphoproliferative lesions in children with AIDS. Pediatr AIDS and HIV Infect: From Fetus to Adolescent 1990;1:44–8.

(22) Soberman N, Leonidas JC, Berdon WE, et al. Parotid enlargement in children seropositive for human immunodeficiency virus: imaging findings. Am J Radiol 1991;157:553–6.

(23) Itescu S. Diffuse infiltrative lymphocytosis syndrome in children and adults infected with HIV-1: a model of rheumatic illness caused by acquired viral infection. Am J Reprod Immunol 1992;28:247–50.

(24) Kontny HU, Sleasman JW, Kingma DW, et al. Multilocular thymic cysts in children with human immunodeficiency virus infection. Clinical and pathological aspects. J Pediatr 1997;131:264–70.

(25) Magrath I, Shad A, Epelman S, et al. Pediatric oncology in countries with limited resources. In: Pizzo PA, Poplack DG, editors. Principles and practice of pediatric oncology. Philadelphia: Lippincott-Raven Publishers, 1997:1395–1420.

(26) Shad A, Magrath IT. Malignant non-Hodgkin's lymphomas in children. In: Pizzo PA, Poplack DG, editors. Principles and practice of pediatric oncology. Philadelphia: Lippincott-Raven Publishers, 1997:545–88.

(27) Guarner J, Brynes RK, Chan WC, et al. Primary non-Hodgkin's lymphoma of the heart in two patients with the acquired immunodeficiency syndrome. Arch Pathol Lab Med 1987;111:254–6.

(28) Siskin GP, Haller JO, Miller S, et al. AIDS-related lymphoma: radiologic features in pediatric patients. Radiology 1995;196:63–6.

(29) Teruya-Feldstein J, Temeck BK, Sloas MM, et al. Pulmonary malignant lymphoma of mucosa-associated lymphoid tissue (MALT) arising in a pediatric HIV-positive patient. Am J Surg Pathol 1995;19:357–63.

(30) Mueller BU, Shad AT, Magrath IT, et al. Malignancies in children with HIV infection. In: Pizzo PA, Wilfert CM, editors. Pediatric AIDS. The challenge of HIV infection in infants, children, and adolescents. Baltimore: Williams & Wilkins, 1994:603–22.

(31) Shad A, Mueller BU, Adde M, et al. Results of a treatment protocol for children with HIV and non-Hodgkin's lymphomas (NHL). J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A54.

(32) Isaacson PG. The MALT lymphoma concept updated. Ann Oncol 1995;6: 319–20.

(33) Slater DN. MALT and SALT: the clue to cutaneous B-cell lymphoproliferative disease. Br J Dermatol 1994;131:557–61.

(34) Connor EM, Marquis J, Oleske JM. Lymphoid interstitial pneumonitis. In: Pizzo PA, Wilfert CM, editors. Pediatric AIDS. The challenge of HIV infection in infants, children, and adolescents. Baltimore: Williams & Wilkins, 1991:343–54.

(35) Pitt J. Lymphocytic interstitial pneumonia. In: Edelson PJ, editor. Childhood AIDS. Philadelphia: W. B. Saunders Co., 1991:89–96.

(36) Scott GB, Hutto C, Makuch RW, et al. Survival in children with perinatally acquired human immunodeficiency virus type 1 infection. N Engl J Med 1989;321:1791–6.

(37) Kingma DW, Weiss A, Sorbora L, et al. Is lymphocytic interstitial pneumonitis an early stage of MALT lymphoma in HIV-infected pediatric patients? J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A54.

(38) Buck BE, Scott GB, Valdes-Dapena M, et al. Kaposi sarcoma in two infants with acquired immune deficiency syndrome. J Pediatr 1993;103: 911–3.

(39) Baum LG, Vinters HV. Lymphadenopathic Kaposi's sarcoma in a pediatric patient with acquired immune deficiency syndrome. Pediatr Pathol 1989; 9:459–65.

(40) Bouquety JC, Siopathis MR, Ravisse PR, et al. Lympho-cutaneous Kaposi's sarcoma in an African pediatric AIDS case. Am J Trop Med Hyg 1989;40:323–5.

(41) Gutierrez-Ortega P, Hierro-Orozco S, Sanchez-Cisneros R, et al. Kaposi's sarcoma in a 6-day-old infant with human immunodeficiency virus. Arch Dermatol 1989;125:432–3.

(42) Connor E, Boccon-Gibod L, Joshi V, et al. Cutaneous acquired immunodeficiency syndrome-associated Kaposi's sarcoma in pediatric patients. Arch Dermatol 1990;126:791–3.

(43) Boshoff C, Schulz TF, Kennedy MM, et al. Kaposi's sarcoma-associated herpesvirus infects endothelial and spindle cells. Nature Med 1995;1: 1274–8.

(44) Gaidano G, Pastore C, Gloghini A, et al. Distribution of human herpesvirus-8 sequences throughout the spectrum of AIDS-related neoplasia. AIDS 1996;10:941–9.

(45) Otsuki T, Kumar S, Ensoli B, et al. Detection of HHV-8/KSHV DNA sequences in AIDS-associated extranodal lymphoid malignancies. Leukemia 1996;10:1358–62.

(46) Strauchen JA, Hauser AD, Burstein D, et al. Body cavity-based malignant lymphoma containing Kaposi sarcoma-associated herpesvirus in an HIV-negative man with previous Kaposi sarcoma. Ann Intern Med 1996;125: 822–5.

(47) Schalling M, Ekman M, Kaaya EE, et al. A role for a new herpes virus (KSHV) in different forms of Kaposi's sarcoma. Nature Med 1995;1: 707–8.

(48) Gao S-J, Kingsley L, Li M, et al. KSHV antibodies among Americans,

Italians and Ugandans with and without Kaposi's sarcoma. Nature Med 1996;2:925–8.
(49) Simpson GR, Schulz TF, Whitby D, et al. Prevalence of Kaposi's sarcoma associated herpesvirus infection measured by antibodies to recombinant capsid protein and latent immunofluorescence antigen. Lancet 996;349: 1133–8.
(50) Monini P, De Lellis L, Fabris M, et al. Kaposi's sarcoma-associated herpesvirus DNA sequences in prostate tissue and human semen. N Engl J Med 1996;334:1168–72.
(51) Lennette ET, Blackbourn DJ, Levy JA. Antibodies to human herpesvirus type 8 in the general population and in Kaposi's sarcoma patients. Lancet 1996;348:858–61.
(52) Balsam D, Segal S. Two smooth muscle tumors in the airway of an HIV-infected child. Pediatr Radiol 1992;22:552–3.
(53) Parola P, Petit N, Azzedine A, et al. Symptomatic leiomyoma of the adrenal gland in a woman with AIDS. AIDS 1996;10:340–1.
(54) Lee ES, Locker J, Nalesnik M, et al. The association of Epstein-Barr virus with smooth muscle tumors occurring after organ transplantation. N Engl J Med 1995;332:19–25.
(55) Melbye M, Cote TR, Kessler L, et al. High incidence of anal cancer among AIDS patients. Lancet 1994;343:636–9.
(56) Palefsky JM. Anal human papillomavirus infection and anal cancer in HIV-positive individuals: an emerging problem. AIDS 1994;8:283–95.
(57) Maiman M, Fruchter RG, Guy L, et al. Human immunodeficiency virus infection and invasive cervical carcinoma. Cancer 1993;71:402–6.
(58) Tweddel G, Heller P, Cunnane M, et al. The correlation between HIV seropositivity, cervical dysplasia, and HPV subtypes 6/11, 16/18, 31/33/35. Gynecol Oncol 1994;52:161–4.
(59) Cripe TP. Human papillomavirus: pediatric perspectives on a family of multifaceted tumorigenic pathogens. Pediatr Infect Dis J 1990;9:836–44.
(60) Tirelli U, Errante D, Dolcetti R, et al. Hodgkin's disease and human immunodeficiency virus infection: clinicopathologic and virologic features of 114 patients from the Italian Cooperative Group on AIDS and Tumors. J Clin Oncol 1995;13:1758–67.
(61) Mazzaro C, Zagonel V, Monfardini S, et al. Hepatitis C and non-Hodgkin's lymphoma. Br J Haematol 1996;94:544–50.
(62) White YS, Johnson PJ, Davison F, et al. Frequency of hepatic HBV-DNA in patients with cirrhosis and hepatocellular carcinoma: relation to serum HBV markers. Br J Cancer 1990;61:909–12.
(63) Pontisso P, Morsica G, Ruvoletto MG, et al. Latent hepatitis B virus infection in childhood hepatocellular carcinoma. Analysis by polymerase chain reaction. Cancer 1992;69:2731–5.
(64) Connor EM, Mofenson LK. Zidovudine for the reduction of perinatal human immunodeficiency virus transmission: Pediatric AIDS Clinical Trials Group protocol 076-results and treatment recommendations. Pediatr Infect Dis J 1995;14:536–41.
(65) Centers for Disease Control and Prevention. Recommendations of the U.S. Public Health Service Task Force on the use of zidovudine to reduce perinatal transmission of human immunodeficiency virus. MMWR Morb Mortal Wkly Rep 1994;43:1–20.
(66) Olivero OA, Anderson LM, Diwan BA, et al. AZT is a genotoxic transplacental carcinogen in animal models. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A29.
(67) Neumann Y, Toren A, Mandel M, et al. Favorable response of pediatric AIDS-related Burkitt's lymphoma treated by aggressive chemotherapy. Med Pediatr Oncol 1993;21:661–4.
(68) Nadal D, Caduff R, Frey E, et al. Non-Hodgkin's lymphoma in four children infected with the human immunodeficiency virus. Cancer 1994;73: 224–30.
(69) Desai N, Miller ST, Rao SP, et al. Malignancies associated with HIV-1 infection in children: one center's experience over 6 years. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A49.

Hodgkin's Disease in the Setting of Human Immunodeficiency Virus Infection

*Alexandra M. Levine**

Although Hodgkin's disease (HD) is not usually associated with congenital or acquired immunodeficiency disorders, recent evidence would suggest a statistically significant increase in HD among individuals infected with human immunodeficiency virus (HIV). In the setting of underlying HIV infection, clinical and pathologic characteristics of HD may differ from usual expectations. Thus, 70%–100% of HIV-infected patients with HD present with systemic "B" symptoms. Likewise, disseminated, stage III or IV disease is reported in approximately 75%–90%. Bone marrow is a common site of extranodal HD, occurring in 40%–50%. Complete response rates after multiagent chemotherapy range from approximately 45% to 70%, although median survival has been only in the range of approximately 18 months. Hematologic toxicity from multiagent chemotherapy may be substantial, even with the use of hematopoietic growth factor support. It is apparent that new strategies of therapeutic intervention must be explored. [Monogr Natl Cancer Inst 1998;23:37–42]

Background: *De Novo* Hodgkin's Disease

Relationship Between Hodgkin's Disease and Underlying Immunodeficiency

Non-Hodgkin's lymphoma is known to occur with increased frequency in various settings of congenital, acquired, or iatrogenic immunosuppression *(1–4)*. It was thus not surprising to observe significantly increased rates of lymphoma in patients infected with human immunodeficiency virus (HIV). Similarly, the other current acquired immunodeficiency syndrome (AIDS)-defining cancers, Kaposi's sarcoma and cervical cancer, are also associated with diverse types of abnormal immunity, including immunosuppression related to organ transplantation or to therapy for autoimmune or malignant disorders *(2,3,5,6)*. Within the context of this background, Hodgkin's disease (HD) represents a somewhat different situation. Thus, while the risk of HD is increased in children with ataxia telangiectasia *(7)*, the magnitude of this increase is quite modest, and HD is not seen with increased frequency in other types of congenital immune deficiency disease. Furthermore, the incidence of HD is not increased in the setting of organ transplantation or in the setting of prior cancer chemotherapy or immunosuppressive therapy for autoimmune disease *(2,3)*. The expectation that HIV-infected patients would be at risk for HD was thus not initially considered a prominent concern.

Epidemiology of *De Novo* HD

In patients who are not infected by HIV, the epidemiology of HD has been well described. Correa, MacMahon, and others *(8–10)* have demonstrated a clear relationship between geographic and socioeconomic factors and the ensuing clinicopathologic characteristics of HD. Thus, the "type 1 pattern" of HD is seen in underdeveloped areas of the world, associated with lower socioeconomic conditions. In this setting, two age peaks of disease are observed, with the first in childhood (primarily affecting boys) and the second in the fifth or sixth decade. The pathologic spectrum of HD in the type 1 setting includes a predominance of mixed-cellularity or lymphocyte-depletion disease, with proclivity toward advanced stage and presence of systemic "B" symptoms. The overall prognosis of such patients is relatively poor, when compared with that of patients in other settings and with other patterns of disease.

The "type 3 pattern" of HD is seen in industrialized nations and is also characterized by two modal peaks of disease *(8–10)*. However, in these areas of higher socioeconomic background, the first modal peak occurs in adolescent or young-adult females, who tend to present with nodular sclerosis HD, often presenting with rather limited stage disease localized to the mediastinal lymph nodes. The second peak occurs in the fifth and sixth decades, is seen primarily in males, and may be associated with a broader spectrum of pathologic subtypes. Various characteristics associated with higher socioeconomic status have been described with increased frequency in young patients who develop HD within type 3 areas.

In the "type 2 pattern," all three age peaks of disease are encountered, with a broad spectrum of pathologic subtypes and clinical outcome. This pattern has been described in developing areas of the world, such as Mexico.

Of interest, areas of lower socioeconomic status within the United States have been described in which the clinicopathologic spectrum of HD resembles that seen in type 2 or developing nations *(11)*. Because of these divergent epidemiologic, clinical, and pathologic characteristics, any differences between HIV-associated and *de novo* HD must be considered within the context of the usual expectations for HD in that particular region under study.

Immune Function in *De Novo* HD

Defects in cell-mediated immunity have been well described in patients with *de novo* HD, occurring even in patients with early stage disease and in those who have been free of HD for many years *(12)*. Thus, patients with HD are often anergic to

**Correspondence to:* Alexandra M. Levine, M.D., Division of Hematology, University of Southern California (USC), USC/Norris Cancer Hospital, 1441 Eastlake Ave., MS 34, Los Angeles, CA 90033.

See "Note" following "References."

© Oxford University Press

delayed cutaneous hypersensitivity testing, as described by Dorothy Reed as early as 1902 [*see (13)*]. While more common in patients with advanced, symptomatic HD, dinitrochlorobenzene skin testing may also be abnormal in patients with good prognosis, early stage HD *(12)*. *In vitro* testing of immune function is also abnormal in patients with *de novo* HD. Thus, decreased mitogenic responses to phytohemagglutinin and to concanavalin A have been described, even prior to the initiation of chemotherapy *(14)*. While significant reductions in T cells have been described in peripheral blood and spleen of approximately 30% of patients with untreated, advanced HD *(15)*, no alteration in the ratio of helper or cytotoxic/suppressor T cells has been found; moreover, there is no evidence of cellular activation *(15)*. Furthermore, most patients with HD do not have lymphocytopenia, indicating that the major cell-mediated defect is a functional impairment, as opposed to an absolute depletion of T cells in blood or tissues.

These defects of cell-mediated immunity may also be surmised by the fact that patients with HD may be at increased risk for certain infections known to be associated with depressed T-cell function. Of interest, increased susceptibility to tuberculosis among patients with HD was well described as early as 1928, while the earliest description of HD was complicated by the fact that some of Thomas Hodgkin's initial HD patients actually had tuberculosis *(16)*. Furthermore, in a large series of 300 adult patients with HD, Notter et al. *(17)* noted the development of opportunistic infections in 4%, consisting of *Pneumocystis carinii* pneumonia, disseminated vaccinia, herpes zoster, and progressive multifocal leukoencephalopathy, among others.

When considering the immune function of patients with HIV-related HD, it is important to realize that any immune defect etiologically related to HIV will, by definition, be augmented by the known defects in cell-mediated immunity that are inherent to HD itself.

Epidemiology of HD in the Setting of HIV Infection

Incidence of HD Among HIV-Infected Persons

Of interest, at the outset of the AIDS epidemic, the incidence of *de novo* HD appeared to have decreased among HIV-negative persons in New York and Los Angeles *(18,19)*. In the evaluation of young, single men in New York City (presumably including HIV-infected individuals or homosexual men at risk for HIV), stationary rates were documented, while an actual decrease in HD among married men in New York State was observed at the same time *(18)*.

In 1992, Hessol et al. *(20)* noted an increase in the standardized morbidity ratio for HD among HIV-infected homosexual men from San Francisco. By use of data from 6704 homosexual men enrolled in the San Francisco City Clinic HIV Cohort Study, cases of HD were ascertained through the computer-matched identification of participants in the population-based Northern California Cancer Center registry. Comparisons were made with rates determined in the general population by use of the Surveillance, Epidemiology, and End Results (SEER)[1] cancer registry. A total of 90 cases of non-Hodgkin's lymphoma were identified, along with eight cases of HD. The age-adjusted standardized morbidity ratio for HD among the HIV-infected men was 5.0 (95% confidence interval [CI] = 2.0–10.3), indicating a statistically significant increase. Furthermore, when these data were compared with those from the SEER registry, the excess risk of HD among HIV-infected homosexual men was 19.3 cases per 100 000 person-years, again indicating a significant increase. While pathologic review was not performed, and while diagnostic misclassification is known to occur, especially in lymphocyte-depleted HD *(21,22)*, the data did indicate that the incidence of HD may have increased in HIV-infected homosexual men *(20)*.

More recently, Lyter et al. *(23)* described an increase in HD among homosexual and bisexual men, who were followed as part of the national Multicenter AIDS Cohort Study (MACS). Thus, within the Pittsburgh MACS cohort, who were followed between 1984 and 1993, two cases of HD were seen among 430 HIV-infected men (with 2344 person-years of follow-up) versus no cases among 769 HIV-negative control subjects (followed for 5708 person-years). While consisting of only two cases, the standardized incidence ratio for seropositives was increased 19.8-fold (95% CI = 2.4–71.5) when compared with that for seronegatives. Furthermore, when these data were compared with those from the population-based SEER registry, the rate of HD among HIV-positive men was 85 per 100 000 person-years versus 4.3 per 100 000 person-years in the general population.

In an attempt to aseertain the full spectrum of malignant disease occurring in association with HIV infection, Goedert and colleagues *(24)* performed a linkage analysis in which all cases of cancer reported in over 98 000 cases of AIDS reported to the Centers for Disease Control and Prevention were linked, case by case, to the population-based cancer registries in the United States. AIDS-related cancers were defined as those that had increased statistically after an initial AIDS diagnosis, with increasing prevalence from 5 years before the diagnosis of AIDS, through 2 years after the initial AIDS diagnosis. In this analysis, the risk of HD was increased 7.6-fold, with the relative risk (RR) statistically increasing from the period prior to AIDS to that occurring after the initial diagnosis of AIDS.

Significant increases in HD have also recently been documented in the National Cancer Institute Hemophilia Cohort (RR = 5.6) *(25)*; in injection drug users with AIDS in New Jersey (odds ratio [OR] = 4.2; 95% CI = 1.4–14.7; P = .04) *(26)*; in HIV-infeeted individuals from Denver, CO *(27)*; and in homosexual men from New South Wales, Australia *(28)*. In the latter report, the RR of HD was 18.3 (95% CI = 8.4–35), with the increase in risk confined to the 2 years immediately before an AIDS diagnosis and the period after AIDS was diagnosed. These data would be consistent with the concept that HD is a late manifestation of HIV disease, which may become even more apparent as HIV-infected individuals live long enough to develop the disease.

HIV-Infected Populations at Risk for HD

Early studies from Europe indicated that HD did occur in HIV-infected individuals, although no formal population-based comparisons were published at that time. However, Monfardini et al. *(29)* from the Italian Cooperative Group for AIDS-related Tumors sent 1962 questionnaires to various professional groups of oncologists, infectious disease specialists, and others, in an

attempt to ascertain potential cases of HD among HIV-infected individuals. A total of 35 cases of HD and 95 cases of lymphoma were identified. Of interest, 89% of the patients with HD had a history of injection drug use, while an additional 9% were homosexual with a history of injection drug use. In contrast, a history of injection drug use was present in 68% of the lymphoma patients, while 5% had a history of both injection drug use and homosexuality. While retrospective and biased by the method of case ascertainment, this study suggested the possibility that injection drug use might be a specific risk factor for HIV-associated HD. Subsequent work from France *(30)* confirmed this suggestion, with 38% of 45 HIV-infected patients with HD providing a history of injection drug use, versus only 12.5% of 168 patients with non-Hodgkin's lymphoma. Roithmann et al. *(31)* confirmed the increased likelihood of HD among injection drug users versus homosexual men from France, while investigators from Spain *(32)* also documented a similar relationship, with 85% of 46 cases of HD occurring in injection drug users and 9% occurring in homosexual men. Furthermore, Ahmed et al. *(33)* also noted that the incidence of HD was three to 10 times higher among prisoners in the United States who were injection drug users, versus those who were not, although the overall incidence of HD had not increased in these prison inmates.

While these earlier studies seemed to indicate an increased proclivity for HD among HIV-infected injection drug users, subsequent studies have not really confirmed this fact. Thus, statistically significant increases in HD have now been described in HIV-infected hemophiliacs *(25)* as well as in HIV-infected homosexual men *(23,28)*. More time will be required to elucidate the true increase in risk of HD among injection drug users, when compared to other HIV-infected populations.

Degree of Immune Deficiency Among HIV-Infected Persons With HD

Relatively little is currently known about the precise level of immune dysfunction in patients with HIV-related HD. Ames et al. *(34)* evaluated 23 such patients, whose median CD4 cell count was 201/mm^3, ranging from 7 to 882/mm^3. In another report published in 1991 from Italy *(35)*, the median CD4 cell count in 38 patients was 254/mm^3 (range, 27–1316/mm^3). A study from France published in 1993 *(30)* described 45 patients with HIV-related HD in whom the median CD4 cell count was 306/mm^3. More recently, CD4 cells were determined at baseline in a group of 21 patients with HIV-related HD *(36)* who were treated prospectively as part of the AIDS Clinical Trials Group (ACTG study #149). The median CD4 cell count was 128/mm^3, ranging from 2 to 972/mm^3. The apparently lower median CD4 cell count seen in this latter series may reflect simple chance alone, or it may be reflective of the fact that this series was accrued later in the epidemic, at a time when patients are living longer, despite falling CD4 cell counts.

Most patients with HIV-related HD have not had an AIDS-defining diagnosis prior to the onset of HD. Thus, in the group of 21 patients studied as part of ACTG study #149 *(36)*, 20% had a history of AIDS prior to HD, consisting of *Mycobacterium avium-intracellulare* in two, *Pneumocystis carinii* pneumonia in two, and Kaposi's sarcoma in one. An additional 50% had symptomatic HIV-related disease prior to the diagnosis of HD, consisting of oral candidiasis, herpes zoster, and/or immune thrombocytopenic purpura. Of interest, although the majority of patients have not had full-blown AIDS prior to HD, many have had a history of reactive lymphadenopathy (persistent, generalized lymphadenopathy), which was reported in 36%–83% of individuals *(30,33,35)*.

Pathologic Aspects of HIV-Associated HD

In *de novo* HD in the United States and Europe, the majority of patients are diagnosed with nodular sclerosis disease, described in 52%–62% of large series *(37)*. Lymphocyte-predominant disease accounts for 8%–21%, while mixed cellularity HD is seen in 24% and lymphocyte depletion subtype in 3%–6% of patients *(37)*. In contrast, the most common type of HD in underdeveloped areas of the world (type 1 pattern) is mixed cellularity, while lymphocyte depletion is the next most frequently diagnosed pathologic subtype of disease *(8–10)*.

Similar to the pathologic spectrum in type 1 areas, patients with HIV-related HD tend to present with either mixed cellularity or lymphocyte depletion HD, diagnosed in 41%–100% of patients *(29,32,34,38–41)*. Serrano et al. *(40)* compared the pathologic characteristics in 22 patients with HIV-related HD with those in 125 uninfected control subjects with HD diagnosed in the same time period, in the same institution in Spain. Mixed cellularity or lymphocyte depletion HD was diagnosed in 68% of the HIV-infected group, versus 35% of the uninfected control subjects. Of interest, when nodular sclerosis was diagnosed in the setting of HIV, it did not occur within the mediastinum, in contrast to HIV-negative patients with nodular sclerosis, in whom mediastinal involvement was documented in 74%.

Clinical Features of HIV-Related HD

Systemic B Symptoms

In general, fever, drenching night sweats, and/or weight loss occur more often in patients with HIV-related HD than in patients with *de novo* disease. Thus, in a large series from the United States or Europe, the prevalence of systemic B symptoms ranged from 30% *(42)* to 62%, the latter described in U.S. patients from minority racial/ethnic backgrounds and lower socioeconomic status *(11)*. In contrast, between 70% and 100% of patients with HIV-related HD have presented with B symptoms *(29,30,32,34,38–41)*. In the series of HIV-positive versus HIV-negative cases of HD from Spain, Serrano et al. *(40)* noted B symptoms in 81% of 22 HIV-infected patients and in 57% of 125 HIV-negative patients.

While thus apparent that fever, night sweats, and/or weight loss may be commonly encountered in the setting of HIV-related HD, it is important to recognize that these symptoms may also be seen in *Mycobacterium avium-intracellulare*, cytomegalovirus disease, and other opportunistic infections. Thus, the assignation of such symptoms to HD must be made only after a careful evaluation has excluded the presence of underlying occult infection.

Clinical Stage of HIV-Related HD at Presentation

In *de novo* HD in the United States or Europe, stage III disease is diagnosed in approximately 27%, while stage IV is documented in 16%–26% *(37)*. In contrast, widely disseminated HD is expected in the majority of patients with HIV-related HD.

Thus, in the study of HIV-positive versus HIV-negative HD in Spain, Serrano et al. *(40)* noted stage III or IV disease in 91% of HIV-infected patients, versus only 46% of those with *de novo* HD. In a similar comparative study from France, Andrieu et al. *(30)* reported stage III or IV disease in 75% of 45 HIV-infected patients and in 33% of 407 HIV-negative individuals. These data would indicate that HD associated with HIV is clearly distinct in terms of the expectation of widely disseminated disease at initial presentation. This is very similar to the expectation in AIDS-related lymphoma, which is distinct from *de novo* lymphomatous disease *(1)*.

The most common site of extranodal HD in HIV-infected patients is the bone marrow, which is involved in approximately 41%–50% of all reported patients *(32,34,39,40,43)*. Of interest, bone marrow involvement may be the first presentation of HIV-related HD, with patients seeking medical attention because of systemic B symptoms and/or signs or symptoms of pancytopenia. The primary diagnosis of HD within bone marrow may be difficult, revealing lymphohistiocytic aggregates and large atypical mononuclear cells that are suggestive but not diagnostic of HD *(44)*. The rather common occurrence of HD within the bone marrow of HIV-infected individuals is in sharp contrast to *de novo* HD, in which Colby et al. *(45)* described such involvement in only 3.5% of 659 patients. While bone marrow involvement is distinctly unusual in *de novo* HD in the United States, when one considers only those patients with mixed cellularity or lymphocyte depletion HD, the prevalence of bone marrow involvement is much higher, at 20%–30%, as reported by O'Carrol et al. *(46)*. Even when one considers only mixed cellularity and lymphocyte depletion disease, however, bone marrow involvement by HD is still more likely in the setting of HIV infection. Thus, Serrano et al. *(40)* described bone marrow involvement by HD in 50% of HIV-positive patients from Spain and in only 10% of 125 HIV-negative patients who were diagnosed and evaluated in the same institution.

Aside from bone marrow, other sites of extranodal HD have included rectum *(34,47)*, tongue *(34)*, and skin *(43,49)*. Involvement of the brain by HD is extremely unusual in *de novo* disease, but it has been described in the setting of HIV *(49)* in an injection drug user who presented with a seizure and was found to have a 3-cm, partially enhancing lesion in the left parietal lobe. Spinal fluid was negative, and the diagnosis of mixed cellularity HD was made after stereotaxic brain biopsy.

More usual sites of extranodal HD have been described in patients with HIV-related disease, but even in this circumstance, the disease is distinct from that expected. Thus, HD in the lung has been described in the absence of mediastinal lymphadenopathy *(43,50)*. Furthermore, hepatic HD has been diagnosed in the absence of splenic involvement *(43)*.

Results of Therapy for HIV-Related HD

The majority of all patients with *de novo* HD experience long-term disease-free survival, with probable cure of disease expected in approximately 70%. When one considers results of radiotherapy for localized stage I or II disease, complete remission is expected in 80%–95%, with 80% of responders in continuous complete remission at 7 years *(51)*. In patients with stage III or IV disease, ABVD chemotherapy (i.e., chemotherapy with doxorubicin, bleomycin, vinblastine, and dacarbazine) is associated with complete remission in 70%–80% of patients, with relapse-free survival in approximately 60%–70% of responders *(52)*.

Results of Retrospective Therapeutic Trials for HIV-Related HD

Prognosis in patients with HIV-related HD is distinctly poorer than that expected in patients with *de novo* disease. Thus, median survival has been in the range of only 8–18 months *(30,32,34,35,40,41,53)*. In the series of HIV-positive versus HIV-negative HD patients reported from Spain *(40)*, 5-year survival was achieved in 80% of 125 HIV-negative patients and in none of 22 HIV-infected patients.

The potential causes of decreased survival in HIV-related HD patients could include early demise as a result of other AIDS-related disorders, decreased efficacy of standard therapy, and/or increased toxicity of treatment. Of interest, while no large prospective series of uniformly treated patients has yet been published, preliminary data would suggest that the results of multiagent chemotherapy may be somewhat less optimal than expected in *de novo* disease. Thus, the comparative study by Serrano et al. *(40)* documented a complete remission in 54% of HIV-infected patients with HD, versus an 84% complete remission rate in those with *de novo* disease. As depicted in Table I, complete remission rates have ranged from 44% to 100% in patients with HIV-related HD after being treated with a variety of regimens *(32,35,38,40,54,55)*.

Prospective Treatment Trials

When one considers prospective therapeutic trials, Errante et al. *(53)* treated 17 patients with epirubicin, vinblastine, and bleomycin. Patients with poor prognostic disease (Eastern Cooperative Oncology Group performance status ≥3 or a history of opportunistic infection) received 50% of planned epirubicin and vinblastine dosing. In addition, these patients received zidovudine at the initiation of chemotherapy, while the others received full-dose chemotherapy with zidovudine initiated only after cycle 3. In patients with good prognosis disease, a complete remission rate of 67% was achieved, while complete remission was achieved in only one (20%) of five patients with poor prog-

Table 1. Hodgkin's disease associated with human immunodeficiency virus: therapeutic results

Regimen(s)*	No. of patients	Complete remission, %	Reference No.
Various	27	44	*(32)*
Various	22	54	*(40)*
Various	32	47	*(35)*
Various	18	83	*(38)*
MOPP or ABVD	8	100	*(55)*
MOPP	22	46	*(54)*
MOPP/ABVD	19	65	
Epirubicin, bleomycin, and vinblastine	29	69	*(56)*
Standard ABVD with G-CSF	21	56	*(36)*

*MOPP = combination chemotherapy with mechlorethamine, vincristine, procarbazine, and prednisone; ABVD = combination chemotherapy with doxorubicin, bleomycin, vinblastine, and dacarbazine; G-CSF = granulocyte colony-stimulating factor.

nosis HIV-related HD. Overall median survival in the 17 patients was 11 months.

Tirelli et al. *(56)* used a similar regimen in 29 patients with HIV-related HD; they employed epirubicin (70 mg/m^2), bleomycin (10 mg/m^2), and vinblastine (6 mg/m^2) given intravenously on day 1, with prednisone (40 mg/m^2) given orally on days 1 through 5. Cycles were given every 21 days. Zidovudine (500 mg/day) was given to all patients who had not taken this drug before, while didanosine was given to the other patients. Granulocyte colony-stimulating factor was given at 5 μg/kg per day subcutaneously from day 6 to day 20 in all cycles. The median CD4 cell count at study entry was 219/mm^3 (range, 6–812/mm^3). Stage III or IV disease was present in 80% of the patients, while 90% had systemic B symptoms. Pathologic evaluation revealed either mixed cellularity or lymphocyte depletion HD in 69% of the patients. The complete remission rate was 69%, with 58% of these patients (i.e., 40% of all patients showing complete remission) remaining disease free at 2 years. Relapse of HD was documented in 30% of patients who had shown complete remission. Opportunistic infections occurred in 28% during therapy. Grade 3 or 4 leukopenia was documented in 34%. Median survival for the group was 14 months. HD was the cause of death in 50% of these patients; HD and opportunistic infection were the causes of death in an additional 5%.

Levine et al. *(36)* have recently reported results from the AIDS Clinical Trials Group, employing standard-dose doxorubicin, bleomycin, vinblastine, and dacarbazine *(52)* in patients with newly diagnosed HIV-related HD. Cycles were repeated every 28 days, and therapy was continued until two cycles beyond complete remission, until six cycles had been given, or until toxicity or progressive disease. Granulocyte colony-stimulating factor was given subcutaneously at a dose of 5 μg/kg per day from days 2 through 14 and days 16 through 28 of each cycle. Antiretroviral agents were withheld during the first two cycles and then initiated with subsequent therapy. Doses of doxorubicin and vinblastine were reduced 75% in the presence of initial bone marrow involvement. A total of 21 patients were accrued, of whom 16 are currently assessable for response. The median age of the patients was 34 years, and 75% were male. The majority (91%) had no history of injection drug use. Median CD4 cell count at study entry was 128/mm^3 (range, 2–972/mm^3), and 20% had a history of AIDS prior to the diagnosis of HD. Systemic B symptoms were present in 86%. Stage IV HD was detected in 62%, while 14% had stage III disease. Sites of extranodal disease included bone marrow in 25%, while kidney, liver, brain, spinal cord, lung, and muscle involvement was also documented. A median of six cycles of ABVD chemotherapy was given *(1–8)*. Complete remission was attained in 56% of the patients, and an additional 25% attained a partial response. Toxicity was significant, with grade 4 neutropenia in 48%, despite the use of granulocyte colony-stimulating factor. Hepatic toxicity was also observed; it was detected in 52% of the patients and consisted of grade 4 transaminasemia in four of 21. Other toxic effects were rather mild. The median survival for the group is 19.5 months, and eight patients have now died, as a result of opportunistic infection in three, liver failure due to reactivation of hepatitis B in one, sepsis in one, bacterial pneumonia in one, and HD in two.

It is apparent from these studies that complete remission may be achieved in the majority of patients with HIV-related HD. However, toxicity from chemotherapy is substantial and is related especially to bone marrow compromise, even when hematopoietic growth factors are routinely employed. Furthermore, the median survival is significantly shorter than that described in patients with *de novo* HD, with the majority of deaths due to AIDS-related complications and other bacterial infections. It is clear that alternative therapeutic strategies must be explored in HIV-infected patients with HD.

References

(1) Levine AM. AIDS-associated malignant lymphoma [review]. Blood 1992; 80:8–20.

(2) Penn I. The occurrence of malignant tumors in immunosuppressed states. Prog Allergy 1986;37:259–300.

(3) Penn I. The incidence of malignancies in transplant recipients. Transplant Proc 1975;7:323–6.

(4) Swinnen LJ, Costanzo-Nordin MR, Fisher SG, O'Sullivan EJ, Johnson MR, Heroux AL, et al. Increased incidence of lymphoproliferative disorder after immunosuppression with the monoclonal antibody OKT3 in cardiac-transplant recipients. N Engl J Med 1990;323:1723–8.

(5) Penn I. Kaposi's sarcoma in organ transplant recipients: report of 20 cases. Transplantation 1979;27:8–11.

(6) Penn I. Depressed immunity and the development of cancer. Clin Exp Immunol 1981;46:459–74.

(7) Spector BD, Perry GS 3d, Kersey JH. Genetically determined immunodeficiency disease (GDID) and malignancy: report from the immunodeficiency—cancer registry. Clin Immunol Immunopathol 1978;11:12–29.

(8) Correa P, O'Conor GT, Berard CW, Axtell LM, Myers MH. International comparability and reproducibility in histologic subclassification of Hodgkin's disease. J Natl Cancer Inst 1973;50:1429–35.

(9) MacMahon B. Epidemiology of Hodgkin's disease. Cancer Res 1966;26: 1189–200.

(10) Grufferman S, Delzell E. Epidemiology of Hodgkin's disease. Epidemiol Rev 1984;6:76–106.

(11) Hu E, Hufford S, Lukes R, Bernstein-Singer M, Sobel G, Gill P, et al. Third-World Hodgkin's disease at the Los Angeles County–University of Southern California Medical Center. J Clin Oncol 1988;6:1285–93.

(12) Eltringham JR, Kaplan HS. Impaired delayed-hypersensitivity responses in 154 patients with untreated Hodgkin's disease. Natl Cancer Inst Monogr 1973;36:107–15.

(13) Kumar RK, Penny R. Cell-mediated immune deficiency in Hodgkin's disease. Immunol Today 1982;3:269–73.

(14) Levy R, Kaplan HS. Impaired lymphocyte function in untreated Hodgkin's disease. N Engl J Med 1974;290:181–6.

(15) Posner MR, Reinherz EL, Breard J, Nadler LM, Rosenthal DS, Schlossman SF. Lymphoid subpopulations of peripheral blood and spleen in untreated Hodgkin's disease. Cancer 1981;48:1170–6.

(16) Ewing J. Neoplastic diseases. Philadelphia: Saunders, 1928.

(17) Notter DT, Grossman PL, Rosenberg SA, Remington JS. Infections in patients with Hodgkin's disease: a clinical study of 300 consecutive adult patients. Rev Infect Disease 1980;2:761–800.

(18) Biggar RJ, Horm J, Goedert JJ, Melbye M. Cancer in a group at risk of acquired immunodeficiency syndrome (AIDS) through 1984. Am J Epidemiol 1987;126:578–86.

(19) Bernstein L, Levin D, Menck H, Ross RK. AIDS-related secular trends in cancer in Los Angeles County men: a comparison by marital status. Cancer Res 1989;49:466–70.

(20) Hessol NA, Katz MH, Liu JY, Buchbinder SP, Rubino CJ, Holmberg SD. Increased incidence of Hodgkin disease in homosexual men with HIV infection. Ann Intern Med 1992;117:309–11.

(21) Glaser SL, Swartz WG. Time trends in Hodgkin's disease incidence. The role of diagnostic accuracy. Cancer 1990;66:2196–204.

(22) Kant JA, Hubbard SM, Longo DL, Simon RM, DeVita VT Jr, Jaffe ES, et al. The pathologic and clinical heterogeneity of lymphocyte-depleted Hodgkin's disease. J Clin Oncol 1986;4:284–94.

(23) Lyter DW, Bryant J, Thackeray R, Rinaldo CR, Kingsley LA. Incidence of human immunodeficiency virus-related and nonrelated malignancies in a large cohort of homosexual men. J Clin Oncol 1995;13:2540–6.

(24) Cote TR, Biggar RJ, Rosenberg PS, DeVesa SS, Percy C, Yellin FJ, et al. Non-Hodgkin's lymphoma among people with AIDS: incidence presentation and public health burden. Int J Cancer 1997;73:645–50.

(25) Rabkin CS, Hilgartner MW, Hedberg KW, Aledort LM, Hatzakis A, Ei-

chinger S, et al. Incidence of lymphomas and other cancers in HIV-infected and HIV-uninfected patients with hemophilia. JAMA 1992;267:1090–4.
(26) Weiss SH, Karcnik T, Patel S, Skurnick JH, Fromowitz F, Schwartz RA, et al. The ten-year cancer and HIV experience of the UMDNJ-University Hospital cancer registry: excess proportions of Hodgkin's lymphoma and lung cancer. J AIDS Human Retrovirol 1997;14:A18.
(27) Myers AM, Toi M, Muldrow M, Rajan G, Kane M, Curiel T, et al. Non AIDS-defining malignancies in HIV positive individuals. J AIDS Human Retrovirol 1997;14:A31.
(28) Grulich A, Wan X, Law M, Coates M, Kaldor J. Rates of non-AIDS defining cancers in people with AIDS. J AIDS Human Retrovirol 1997; 14:A18.
(29) Monfardini S, Tirelli U, Vaccher E, Ambrosini A, Andriani A, Silvestroni IB, et al. Malignant lymphomas in patients with or at risk for AIDS: a report of 50 cases observed in Italy. Cancer Detect Prev 1988;12:237–41.
(30) Andrieu JM, Roithmann S, Tourani JM, Levy R, Desablens B, le Maignan C, et al. Hodgkin's disease during HIV1 infection: the French registry experience. French registry of HIV-associated tumors. Ann Oncol 1993;4: 635–41.
(31) Roithmann S, Tourani JM, Andrieu JM. Hodgkin's disease in HIV-infected intravenous drug abusers. N Engl J Med 1990;323:275–6.
(32) Rubio R. Hodgkin's disease associated with human immunodeficiency virus infection. A clinical study of 46 cases. Cooperative Study Group of Malignancies Associated with HIV Infection of Madrid. Cancer 1994;73: 2400–7.
(33) Ahmed T, Wormser GP, Stahl RE, Mamtani R, Cimino J, Glasser M, et al. Malignant lymphomas in a population at risk for acquired immune deficiency syndrome. Cancer 1987;60:719–23.
(34) Ames ED, Conjalka MS, Goldberg AF, Hirschman R, Jain S, Distenfeld A, et al. Hodgkin's disease and AIDS. Twenty-three new cases and a review of the literature. Hematol Oncol Clin North Am 1991;5:343–56.
(35) Monfardini S, Tirelli U, Vaccher E, Foa R, Gavosto F. Hodgkin's disease in 63 intravenous drug users infected with human immunodeficiency virus. Gruppo Italiano Cooperativo AIDS & Tumori (GICAT). Ann Oncol 1991;2(suppl 2):201–5.
(36) Levine AM, Cheung T, Tulpule A, Huang J, Testa M. Preliminary results of AIDS Clinical Trials Group (ACTG) study #149: phase II trial of ABVD chemotherapy with G-CSF in HIV-infected patients with Hodgkin's disease. J AIDS Human Retrovirol 1997;14:A12.
(37) Davis S, Dahlberg S, Myers MH, Chen A, Steinhorn SC. Hodgkin's disease in the United States: a comparison of patient characteristics and survival in the Centralized Cancer Patient Data System and the Surveillance, Epidemiology, and End Results Program. J Natl Cancer Inst 1987;78:471–8.
(38) Gold JE, Altarac D, Ree HJ, Khan A, Sordillo PP, Zalusky R. HIV-associated Hodgkin disease: a clinical study of 18 cases and review of the literature. Am J Hematol 1991;36:93–9.
(39) Ree HJ, Strauchen JA, Khan AA, Gold JE, Crowley JP, Kahn H, et al. Human immunodeficiency virus-associated Hodgkin's disease. Clinicopathologic studies of 24 cases and preponderance of mixed cellularity type characterized by the occurrence of fibrohistiocytoid stromal cells. Cancer 1991;67:1614–21.
(40) Serrano M, Bellas C, Campo E, Ribera J, Martin C, Rubio R, et al. Hodgkin's disease in patients with antibodies to human immunodeficiency virus. A study of 22 patients. Cancer 1990;65:2248–54.
(41) Tirelli U, Errante D, Vaccher E, Repetto L, Rizzardini G, Spina M, et al. Hodgkin's disease in 92 patients with HIV infection: the Italian experience. GICAT (Italian Cooperative Group on AIDS & Tumors). Ann Oncol 1992;3(suppl 4):69–72.
(42) Kaplan HS. Hodgkin's disease. 2nd ed. Cambridge, MA: Harvard University Press, 1980.
(43) Schoeppel SL, Hoppe RT, Dorfman RF, Horning SJ, Chew TG, Weiss LM, et al. AIDS and Hodgkin's disease. Front Radiat Ther Oncol 1985;19: 66–73.
(44) Karcher DS. Clinically unsuspected Hodgkin disease presenting initially in the bone marrow of patients infected with the human immunodeficiency virus. Cancer 1993;71:1235–8.
(45) Colby TV, Hoppe RT, Warnke RA. Hodgkin's disease at autopsy: 1972–1977. Cancer 1982;49:1848–58.
(46) O'Carrol DI, McKenna RW, Brunning RD. Bone marrow manifestations of Hodgkin's disease. Cancer 1976;38:1717–28.
(47) Picard O, de Gramont A, Krulik M, Gonzalez G, Debray J, Cady J. Rectal Hodgkin disease and the acquired immunodeficiency syndrome. Ann Intern Med 1987;106:775.
(48) Shaw MT, Jacobs SR. Cutaneous Hodgkin's disease in a patient with human immunodeficiency virus infection. Cancer 1989;64:2585–7.
(49) Hair LS, Rogers JD, Chadburn A, Sisti MB, Knowles DM, Powers JM. Intracerebral Hodgkin's disease in a human immunodeficiency virus-seropositive patient. Cancer 1991;67:2931–4.
(50) Schieb RG, Siegel RS. Atypical Hodgkin's disease and the acquired immunodeficiency syndrome. Ann Intern Med 1985;102:554.
(51) Hoppe RT, Hanlon AL, Hanks GE, Owen JB. Progress in the treatment of Hodgkin's disease in the United States, 1973 versus 1983. The Patterns of Care Study. Cancer 1994;74:3198–203.
(52) Canellos GP, Anderson JR, Propert KJ, Nissen N, Cooper MR, Henderson ES, et al. Chemotherapy of advanced Hodgkin's disease with MOPP, ABVD, or MOPP alternating with ABVD. N Engl J Med 1992;327: 1478–84.
(53) Errante D, Tirelli U, Gastaldi R, Milo D, Nosari AM, Rossi G, et al. Combined antineoplastic and antiretroviral therapy for patients with Hodgkin's disease and human immunodeficiency virus infection. A prospective study of 17 patients. The Italian Cooperative Group on AIDS and Tumors (GICAT). Cancer 1994;73:437–44.
(54) Tirelli U, Vaccher E, Rezza G, Broccia G, Fassio PG, Gobbi M, et al. Hodgkin disease and infection with the human immunodeficiency virus in Italy. Ann Intern Med 1988;108:309–10.
(55) Newcom SR, Ward M, Napoli VM, Kutner M. Treatment of human immunodeficiency virus-associated Hodgkin disease. Is there a clue regarding the cause of Hodgkin disease? Cancer 1993;71:3138–45.
(56) Tirelli U, Errante D, Santarossa S, Spina M, Gisselbrecht C, Marolleau P, et al. Epirubicin, bleomycin, vinblastine, and prednisone chemotherapy in combination with antiretroviral therapy and primary use of G-CSF for patients with Hodgkin's disease and HIV infection. J AIDS and Human Retrovirol 1997;14:A32.

Note

[1]*Editor's note:* SEER is a set of geographically defined, population-based central tumor registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Each registry annually submits its cases to the NCI on a computer tape. These computer tapes are then edited by the NCI and made available for analysis.

Management of Cervical Neoplasia in Human Immunodeficiency Virus-Infected Women

*Mitchell Maiman**

The existence of cervical neoplasia in women with human immunodeficiency virus (HIV) represents one of the most serious challenges in the oncologic care of immunosuppressed patients. While the development of most cancers in the immunosuppressed patient can be attributed solely to immune deficiency, the relationship between squamous cell neoplasia of the cervix and HIV is quite unique because of common sexual behavioral risk factors. Screening strategies in HIV-positive women must take into account the high prevalence of cervical dysplasia in this subgroup as well as the limitations of cytologic screening. Cervical dysplasia in HIV-positive women may be of higher grade than in HIV-negative patients, with more extensive involvement of the lower genital tract with HPV-associated lesions. The presence and severity of cervical neoplasia in HIV-positive women correlate with both quantitative and qualitative T-cell function. Standard therapies for preinvasive cervical disease have yielded suboptimal results with high recurrence rates. While poor treatment results of standard ablative and excisional therapies warrant unique therapeutic strategies, one must recognize that close surveillance and repetitive treatment have been successful in preventing progressive neoplasia and invasive cervical carcinoma. The disease characteristics of invasive cervical carcinoma may take a more aggressive clinical course in HIV-infected women. HIV-positive women with cervical cancer have higher recurrence and death rates with shorter intervals to recurrence and death than do HIV-negative control subjects. CD4 status does influence subsequent outcome. In general, the same principles that guide the oncologic management of cervical cancer in immunocompetent patients should be applied. However, extremely close monitoring for both therapeutic efficacy and unusual toxicity must be instituted. [Monogr Natl Cancer Inst 1998;23:43–49]

Human immunodeficiency virus (HIV) infection continues to be a national and international health problem of epidemic proportions. Despite the fact that new acquired immunodeficiency syndrome (AIDS) cases increased by less than 5% for the fifth year in a row, a number well below the rate of increase in the epidemic's first decade, the incidence of AIDS cases among women continues to increase, particularly in minority women. Approximately 19% of new adult and adolescent cases of AIDS in the United States last year were in women, and women represent the subgroup with the greatest rate of increase compared with any other defined population in North America. As with cervical neoplasia, HIV infection is largely a disease of women in their reproductive years, with the incidence of both diseases significantly higher in women of color.

The existence of cervical neoplasia in women infected with HIV represents one of the most serious challenges in the oncologic care of immunosuppressed patients. While the development of most illnesses and cancers in HIV-infected patients can largely be attributed to immune deficiency, the relationship between cervical neoplasia and HIV infection is quite unique. Both cervical carcinoma and HIV infection are, in part, sexually transmitted diseases, with oncogenic types of human papillomavirus (HPV) infection the implicated viral carcinogen associated with cervical cancer. Therefore, an association between cervical cancer and HIV can be anticipated not only on the basis of immunosuppression but also because of shared common sexual behavioral risk factors. Thus, while immunosuppressed women, such as renal transplant patients receiving highly immunosuppressive drugs, are known to be at high risk for lower genital tract neoplasia, immunodeficient HIV-infected women are perhaps the highest risk subgroup that we know of. While these factors likely play the major role in the pathogenesis of cervical neoplasia in HIV-infected women, direct interactions between HIV and HPV at the molecular level, the effects of HIV on the local mucosal immune response, enhancement of HPV regulatory expression by the HIV-1 tat protein, and HIV-induced perturbations of paracrine or autocrine factors that influence HPV gene expression must also be considered.

Screening Issues

Because a well-defined precursor lesion for cervical cancer can be detected by screening, organized screening programs for cervical neoplasia can be expected to reduce both the incidence and the mortality rates. Since the prevalence of disease is far greater in HIV-positive patients than in the general population, optimal screening strategies take on an increased importance. HIV-positive women have as much as a 10-fold increased rate of abnormal cytology, including a wide range of cellular and inflammatory changes, such as hyperkeratosis, parakeratosis, trichonomas, herpes, inflammatory atypia, HPV-related changes, and varying degrees of cervical neoplasia *(1)*. Higher rates of abnormalities have also been demonstrated after seroconversion than before seroconversion. Studies from several centers have demonstrated that women who are immunodeficient from HIV infection have cytologic abnormality rates of between 30% and 60% *(2–4)* and Pap smears consistent with

**Correspondence to:* Mitchell Maiman, M.D., Department of Gynecologic Oncology, State University of New York-Health Science Center at Brooklyn, 450 Clarkson Ave., Box 24, Brooklyn, NY 11203.

© Oxford University Press

cervical dysplasia from 15% to 40% *(5)*. Most studies also consistently demonstrate that the prevalence of such abnormalities increases as immunodeficiency becomes more severe.

Screening strategies in HIV-positive women must take into account the high prevalence of cervical dysplasia in this subgroup as well as the limitations of cytologic screening, the relatively high noncompliance rate, and the possibility of accelerated progression of disease. These factors make initial accurate diagnosis critical. Although the Centers for Disease Control and Prevention (CDC) continues to recommend Pap smears as the sole screening method, many authors have advocated the need for baseline colposcopy in HIV-infected women because of the inaccuracy of cytology in consistently predicting histology, low negative predictive values of the Pap smear, decreased sensitivity, and discordances between cytologic smear and biopsy results *(6–10)*. Other authors have found that HIV-infected women had significantly more smears of limited adequacy due to obscuring blood or inflammation and higher rates of concomitant vaginal infections leading to under-read smears *(11)*. Maiman et al. *(12)* in a unique study of 248 HIV-infected women, all of whom had cytology, colposcopy, and biopsy, found that 38% of all cervical intraepithelial neoplasia (CIN) in 13% of the total patients would have been missed if routine colposcopy and biopsy were not performed. Interestingly, similar high false-negative rates of cytology have been reported in immunosuppressed women after renal transplantation *(13)*.

The limitations of cytologic screening become more glaring as the prevalence of cervical dysplasia increases in a given population. Although more frequent cytologic screening (every 6 months) is to be advocated, a reasonable but not universally accepted approach that we use at our institution involves baseline colposcopy or cervicography in HIV-positive women once the diagnosis is made (Fig. 1). Patients with normal colposcopy could then undergo more aggressive cytologic screening than the normal population, while those diagnosed and treated for CIN undergo colposcopy with liberal biopsy every 4–6 months for 2 years and seminannual Pap tests thereafter. Alternatively, screening strategies may be based on baseline immune status, with more aggressive techniques reserved for patients with CD4 counts of less than 500/mm^3. Of course, all strategies must take into account availability of colposcopic resources and individualized knowledge of risk in given patient populations.

In contrast to the inherent lack of sensitivity of normal cervical cytology, abnormal cytology is extremely accurate in predicting CIN on histology, and Pap smears indicating CIN must be taken very seriously *(14)*. Specificities of 84% or greater have been reported with even higher predictive values *(10)*. Therefore, the vast majority of such patients will indeed have cervical pathology, justifying immediate treatment during initial colposcopic evaluation of a significantly abnormal Pap smear. This "see-and-treat" approach using excisional methods such as the loop electrosurgical excision procedure may be particularly appropriate for the HIV-infected woman.

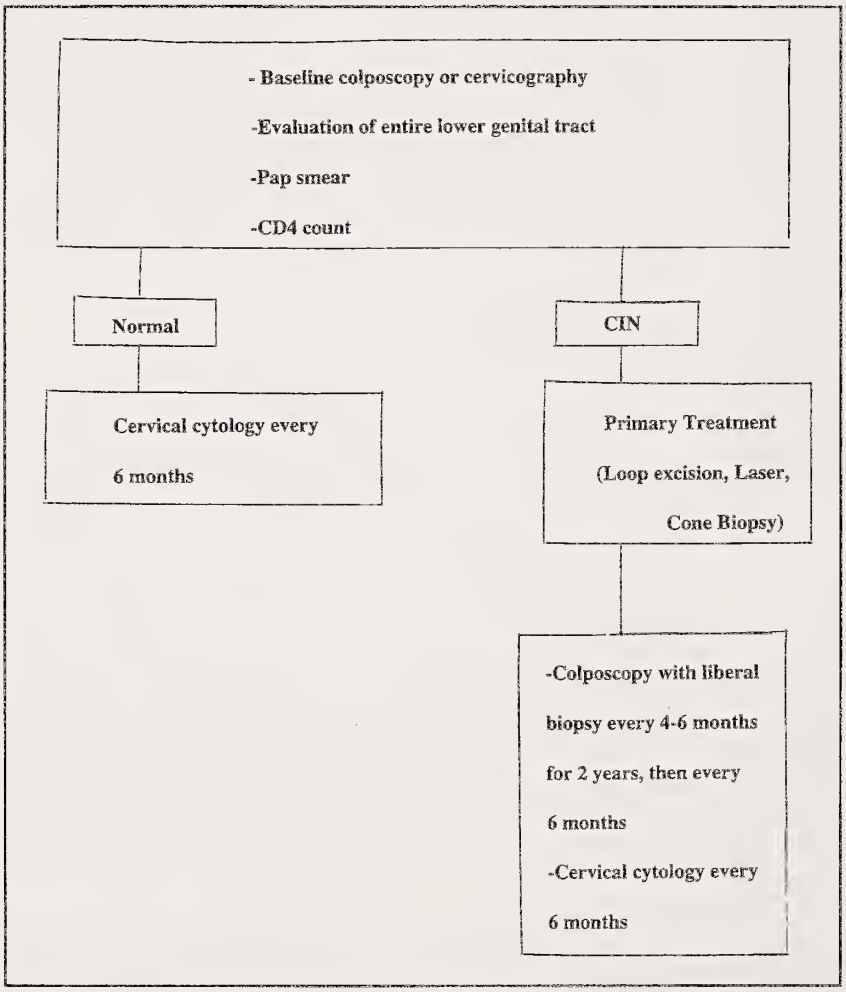


Fig. 1. Screening for cervical intraepithelial neoplasia in human immunodeficiency virus-positive women.

The prevalence of HIV infection in women with invasive cervical carcinoma may be even higher than in those with preinvasive disease. Maiman et al. *(15)* reported a 19% seropositivity rate in women under the age of 50 years in a high-risk population in Brooklyn. Of course, this high prevalence rate is, in fact, reflective of the high HIV infection rate in our community and would be expectedly lower in lower risk geographic areas. Most important, the majority of these women were asymptomatic with regard to HIV disease and die of cervical cancer, not AIDS; therefore, only HIV screening programs would have detected their positive serostatus. At our institution, we currently recommend HIV counseling and testing in all younger (<50 years of age) patients with cervical cancer, since test results may have a significant impact on oncologic therapeutic strategies. The Gynecologic Oncology Group (GOG) is currently conducting a nationwide screening study involving HIV testing and follow-up of newly diagnosed patients with invasive cervical carcinoma.

Invasive Cervical Carcinoma

On January 1, 1993, the CDC expanded the surveillance case definition of AIDS to include

HIV-positive women with invasive cervical cancer. As an AIDS-defining illness, it is required that physicians and hospitals report cases of cervical cancer in HIV-infected women to their local health departments. These changes have served to educate the health care community concerning this relationship and stimulate more aggressive testing programs. In the first year of the expanded definition, approximately 1.3% of women with AIDS-defining illnesses 13 years of age or older had cervical cancer *(16)*. However, HIV testing in women with cervical cancer was far from routine, and documentation of AIDS cases varied considerably based on the reporting practices of physicians and hospitals. Other factors may also contribute to the underdiagnosis of cervical cancer in HIV-infected women, including lack of HIV testing in older women, poor access to health care, and the coexistence of more acute and life-threatening AIDS-defining opportunistic infections that inhibit both the diagnosis and the reporting of cervical cancer. In an urban population at high risk for both diseases in Brooklyn, where a routine HIV testing program was instituted in all cervical cancer patients 50 years of age or younger *(17)*, cervical cancer was the sixth most common AIDS-defining illness in women representing 4% of the subjects as well as the most common AIDS-related cancer in women (55%), followed by lymphoma (29%) and Kaposi's sarcoma (16%).

The signs and symptoms of cervical cancer may take on typical or atypical presentations in HIV-infected women. Ideally, patients should be diagnosed by classic screening methods of cytology, colposcopy, and biopsy if disease is to be detected in the preinvasive or early invasive phase. When patients present with symptoms, more advanced disease is often found, and vaginal bleeding or postcoital bleeding is most commonly reported. Malodorous vaginal discharge is also quite common, as is pelvic pain, back pain, or lower abdominal pain. Leg pain, edema, weight loss, or obstructive uropathy is indicative of more advanced disease. In HIV-positive patients, metastatic disease may occur in both common and uncommon sites *(18)*, and unusual extracervical metastases have been described in the psoas muscles, periclitoral area, and spinal cord as well as malignant ascites *(19,20)*. Diagnosis can be even more difficult, because the classic signs of systemic cancer may mimic the subtle manifestations of HIV disease. Low-grade fevers, unexplained weight loss, gastrointestinal disturbances, and fatigue may occur in both disease processes. Lymphadenopathy, either clinically detected in the left supraclavicular (scalene) or inguinal nodes, or retroperitoneal (pelvic or para-aortic) nodes, discovered at the time of surgery or radiologic imaging, must not be assumed to be metastatic cancer; HIV-infected patients will often have very large, suspicious nodes that may be secondary to follicular hyperplasia, not to a tumor. This pathologic diagnosis, which includes mononuclear cell proliferation, polykaryoctes, epithelioid histiocytes, and mantle-zone loss, while not pathognomonic for HIV, is highly suggestive.

In addition, coexistent pelvic infection may present diagnostic and therapeutic dilemmas. HIV-infected women have high rates of concomitant pelvic inflammatory disease that is more often refractory to antibiotic therapy. Pelvic abscesses may mimic metastatic cervical cancer and pelvic infection may contribute to the development and spread of disease and increase the failure rate and morbidity of therapeutic interventions.

Cervical cancer is the only gynecologic cancer that is clinically staged, which is assigned by the current staging system of the International Federation of Gynecology and Obstetrics (FIGO). Once a clinical stage is assigned and treatment has been initiated, the stage must not be changed because of subsequent findings by either extended clinical or surgical staging. Bimanual rectovaginal pelvic examination, performed under anesthesia if necessary, is the most important part of clinical staging, and only certain additional studies for extracervical disease are allowed by FIGO, including intravenous pyelogram, barium enema, chest and skeletal x ray, cystoscopy, and sigmoidoscopy. However, other more practical and useful studies are often employed that provide more specific information, including computerized axial tomography, magnetic resonance imaging, and laparoscopic lymph node biopsies or extraperitoneal lymph node biopsies by laparotomy. Tumor markers, specifically squamous cell carcinoma antigen or CA 125 for the respective cell types of squamous or adenocarcinoma, are useful parameters to follow patients with during and after treatment.

Cancers of the cervix most often spread by direct invasion into the cervical stroma and laterally into the parametrial tissues as well as into the vagina and corpus uteri. Lymphatic metastases are also quite common and are involved in a progressive fashion beginning with the pelvic (obturator, external and internal iliac) nodes and then involving the para-aortic, inguinal, and scalene nodes. The incidence of positive pelvic and para-aortic nodes ranges from 15% to 5%, respectively, for patients with stage IB disease to 45% and 30% for patients with stage III. However, HIV-positive patients may have higher grade tumors and higher rates of lymph node involvement than expected by stage alone. Cervical cancer may also spread by blood-borne metastases or intraperitoneal implantation or present or recur with painful bony metastases.

Although much of the data, thus far, has been limited to case reports and single institution studies, it seems apparent that in some HIV-infected women, the disease characteristics may take a more aggressive clinical course *(15,21)*. This parallels the examples observed in other AIDS-related cancers. HIV-positive women present with more advanced disease than do HIV-negative women with cervical cancer. In the largest study to date at the Health Science Center at Brooklyn comparing 16 seropositive to 68 seronegative women, significantly more advanced disease was found in the HIV-positive women *(15)*. Half of the seropositive patients presented with stage III or IV disease, compared with 19% in seronegative controls, and only one infected patient remained with early disease after careful surgical–pathologic staging. Almost 70% of the HIV-infected women had stage III or higher surgical–pathologic stage compared with 28% of seronegative controls. HIV-infected women with cervical cancer may be quite young, such as the case of a 16-year-old who presented with stage IIIB disease and died at age 17 years *(18)*. Most HIV-positive patients can be expected to have squamous cell carcinomas as was found in the above study where 15 of 16 patients had squamous cell cancers, with the remaining patient having an adenosquamous tumor *(15)*. Most important, the majority of patients die of cervical cancer before they died of AIDS and are usually asymptomatic with regard to HIV infection. Therefore, only HIV testing programs of such patients will

enable the physician to make the diagnosis and realize the full impact of the interaction of these two disease processes.

The HIV-positive women with cervical cancer have higher recurrence and death rates with shorter intervals to recurrence and death than do HIV-negative control subjects *(15)*. Mean intervals to recurrence are extremely short, and many patients retain persistent disease after primary treatment. Median time to death was found to be 10 months in seropositive women compared with 23 months in seronegative patients. Several other case reports have described examples of rapidly progressive cervical cancer in such women. As with preinvasive disease, the relationship between immune function and disease status is apparent. The mean absolute CD4 count in such patients was 360/mm^3 compared with 830/mm^3 in uninfected women. One can infer from these values that the typical HIV-infected woman with cervical cancer would not meet the immunologic definition for AIDS solely by CD4 count (<200/mm^3), emphasizing the importance of the recent inclusion of seropositive women with cervical cancer as AIDS patients. Although stage of cancer may not predict CD4 levels, immune status does influence subsequent outcome. Patients with counts greater than 500/mm^3 have had more favorable disease courses; therefore, management decisions in HIV-infected women with cervical cancer should carefully consider pretreatment immune function, since positive serostatus alone may not necessarily and uniformly confer a dismal outcome *(15)*. HIV-related immunodeficiency may contribute heavily to the natural history of cervical carcinoma, and HIV-positive patients need not demonstrate other signs of immunosuppression such as opportunistic infections for their neoplasia to be adversely affected by HIV.

The characteristics of HIV disease in cervical cancer may be different in cervical cancer patients when compared with HIV-positive patients with other AIDS-related cancers. Women with invasive cervical cancer are less immunosuppressed than women with other AIDS opportunistic illnesses and may be expected to have CD4 counts about twice as high as those with Kaposi's sarcoma and non-Hodgkin's lymphoma (312–153/mm^3) *(17)*. In the majority of women with cervical cancer, HIV infection was diagnosed at the time of cancer presentation, whereas in women with other cancers, HIV diagnosis preceded cancer diagnosis by a mean of 2.7 years. Although the interval from cancer diagnosis to death may be similar in all AIDS-related cancers, cancer was found to be the cause of death more often with cervical cancer patients (95%) compared with those with other cancers (60%). Although patients with cervical cancer as their AIDS-defining illness may be slightly younger than those women without cervical cancer, the distribution of mode of HIV transmission (heterosexual versus injection drug abuse) and race (black versus Hispanic versus white) was found to be remarkably similar *(16)*.

The management of HIV-positive patients with cervical cancer is among the most challenging tasks faced by the oncologic team. In general, the same principles that guide the oncologic management of cervical cancer in immunocompetent patients should be applied (Fig. 2). However, extremely close monitoring for both therapeutic efficacy and unusual toxicity must be instituted.

Radical hysterectomy and pelvic lymphadenectomy should be performed in most cases of stage IA2 and IB1 and in some cases of stage IB2 and IIA cervical cancer where cervical size is not too enlarged and vaginal involvement is minimal. Radical oncologic surgery in HIV-positive women can be performed for the usual indications, and surgical decisions should be based on oncologic issues not HIV status. As has been demonstrated in other types of operations in the HIV-positive patients, women with relatively good immune function tolerate surgery well with no significant excess morbidity. Prophylactic antibiotics should be routinely used, and standard surgical precautions should be taken to prevent surgical transmission. The transmission rate of HIV from patient to health care worker is extremely low, estimated to be about 1 in 320.

Although radiation therapy can be used in all stages of cervical cancer and has identical cure rates when compared with radical surgery in stage IB, the issues of ovarian conservation and better vaginal sexual function make radical hysterectomy the preferred modality in younger patients. Ovarian transposition should be considered at the time of surgery if postoperative pelvic radiation is a strong possibility. One may expeet a urologic fistula rate of 1%–2% and a chronic bladder atony rate of 3% after radical hysterectomy compared with an intestinal and urinary stricture and fistula rate of 3%–5% and rate of chronic radiation fibrosis of the bowel or bladder of 6%–8% with pelvic radiation therapy. Another advantage of surgery over radiation is the availability of surgical–pathologic data such as lymph node status for which postoperative therapy can then be designed.

Since most HIV-infected patients with cervical cancer present with more advanced disease, radiation therapy is usually the cornerstone of treatment and is the basis for therapy in stages II–IV. Pelvic radiation therapy begins with external-beam therapy designed to shrink the primary tumor and create better geometry for the brachytherapy insertions to follow. Pelvic fields are usually 15 × 5 cm, extending to a 2-cm margin lateral

Stage Ia1:	**Cold knife therapeutic cone biopsy if fertility desired; otherwise, simple hysterectomy**
Stage IA2 **Stage IB1**	**Radical hysterectomy with pelvic lymphadenectomy Alternatively, radiation therapy in poor surgical candidates**
Stage IB2 **Stage IIA**	**Radiation therapy +/- simple hysterectomy; <u>or</u> Radical hysterectomy with pelvic lymphadenectomy; <u>or</u> Neoadjuvant chemotherapy + radical surgery**
Stage IIB-IVA:	**Radiation therapy +/- chemosensitization**
Stage IVB:	**Chemotherapy +/- radiation therapy**
Recurrent Disease:	**Pelvic Exenteration (central disease), otherwise Palliative Chemotherapy**

Fig. 2. Treatment recommendations for cervical carcinoma in human immunodeficiency virus-infected women.

to the bony pelvis and inferiorly to the border of the obturator foramen. The superior margin can be extended to 18 cm to cover the common iliac nodes or even higher if para-aortic lymph node metastases are evident. The usual dosages of pelvic radiotherapy delivered include 7000–8000 cGy to point A and 6000 cGy to point B. Some HIV-positive patients may respond poorly to radiation therapy alone and regimens that incorporate radiation sensitizers that are being used more frequently in general should be strongly considered. The two most common sensitizing regimens include cisplatin (50 mg/m^2, weeks 1 and 6) combined with fluorouracil (1000 mg/m^2, continuous infusion × 4 days, weeks 1 and 6) or oral hydroxyurea (3 g/m^2, twice per week × 5 weeks). One must also keep in mind that pelvic radiation is associated with transient lymphopenia and depressed T-cell function, which may lead to further immunologic compromise in the HIV-positive patient, and that anecdotal reports of poor tolerance and increased morbidity from pelvic radiation therapy have been made, although adequate studies have not yet been performed to evaluate this issue.

Chemotherapy should be used in cases of systemic disease (pulmonary or liver metastases) or recurrent disease after radiation failures in those patients not eligible for pelvic exenteration. However, recurrent cervical cancer in this setting is not considered curable with chemotherapy, and treatment is palliative. Regimens that incorporate drugs that are both active in cervical carcinoma and relatively bone marrow sparing should be used, with close monitoring of all hematologic indices. Cisplatin (50–75 mg/m^2) is the drug of choice and may be combined with bleomycin (20 U/m^2; maximum, 30 U) and vincristine (1 mg/m^2). Recently, neoadjuvant chemotherapy has been used in patients with stage IB2 or stage II cervical cancer with similar agents followed by radical hysterectomy with good results; however, the efficacy of this approach in HIV-positive patients is unknown.

Other novel therapeutic approaches, such as the use of interferons, retinoids, bone marrow support, or vaccine therapy, may represent future investigative treatment options. The Gynecology Oncology Group is presently evaluating the role of interferon alfa and isotretinoin with or without antiretroviral therapy in the neoadjuvant or advanced/recurrent setting in the treatment of HIV-infected women with cervical cancer. Innovative treatment regimens and research protocols such as these are needed to combat the interaction of these two potentially life-threatening disease processes.

Last, the treatment of HIV and its sequelae with anti-HIV-1 agents and prophylactic regimens is exceedingly complex and always changing. Combination therapy with nucleoside analogues and protease inhibitors is now standard and initiated earlier in the course of HIV disease. Careful attention to overlapping side effects of these regimens when combined with oncologic therapy for cervical cancer as well as potential synergistic antineoplastic effects will be an important area of future investigation.

Preinvasive Cervical Neoplasia

HIV-seropositive women represent perhaps the highest risk group encountered for the development of CIN. In the most recent classification system of HIV-related diseases, the CDC has identified moderate to severe cervical dysplasia and carcinoma *in situ* as category B conditions. Independent studies *(2,4,6,10)* have estimated the prevalence of CIN in this subgroup to be between 20% and 50%. Maiman et al. *(12)*, in a study of HIV-positive women without AIDS-defining illness, found the prevalence of CIN on histology to be 32% in a cohort of 248 patients. Cervical dysplasia of HIV-positive women may be of higher grade than in seronegative women, with more extensive involvement of the lower genital tract with HPV-associated lesions *(18)*. Extensive cervical involvement, endocervical involvement, and multisite (vagina, vulva, and perianal) disease are more common. Natural history studies examining the biologic behavior of cervical HPV and CIN strongly suggest that disease is more aggressive in HIV-positive patients. Conti *(22)* found fourfold higher progressions and threefold lower regression rates of untreated HPV-related cervical lesions in infected women compared with HIV-negative controls, and Petry et al. *(23)* found only a 27% regression rate of CIN 1 lesions in immunosuppressed HIV-positive and transplant patients compared with 62% in immunocompetent control subjects. Higher rates of more oncogenic HPV subtype infection, multiple-type HPV infection, and unspecified-type HPV infection have been reported by many investigators and may help explain the more aggressive cervical pathology that develops in HIV-infected women *(12,24–26)*.

Numerous studies *(6,9,27)* have demonstrated the relationship between HIV-associated immunosuppression and the development of CIN, and the presence and severity of cervical neoplasia are associated with quantitative T-cell function. In one study *(6)*, HIV-positive patients with CIN had absolute CD4 counts and T4:T8 ratios roughly half those of HIV-positive patients without CIN, and patients with AIDS-defining illness are more likely to have cervical disease than are asymptomatic HIV-positive patients. Schafer et al. *(27)* and Wright et al. *(28)* concluded that a CD4 lymphocyte count of 200/mm^3 was independently associated with CIN. The concept of worsening immunodeficiency increasing the risk of cervical pathology may be used to individualize screening and surveillance strategies in HIV-positive women.

The treatment of preinvasive cervical disease in HIV-infected women is among the most challenging and frustrating tasks faced by the practicing gynecologist. In general, standard therapeutic strategies for immunocompetent women apply, but the increased risk for treatment failures and chronic nature of disease in such women is well documented.

Excisional methods, which include LEEP (loop electrosurgical excision procedure), laser cone, and cold knife cone biopsy, are preferred over ablative methods, such as cryotherapy and laser vaporization. Excisional methods have the advantage of confirmation of histology and documentation of negative margins, which is of particular importance in HIV-positive women in whom disease may be more extensive. Del Priore et al. *(29)* reported that colposcopically directed biopsies may be poor predictors of histology on excisional cone specimens in HIV-seropositive women, since 47% with CIN II–III on cone biopsy had only CIN I or HPV on punch biopsy compared with only 9% in HIV-seronegative patients Additionally, with advances in the technology in electrosurgical generators and the development of large wire loops with insulated bases, LEEP has allowed excision of the cervical transformation zone and distal canal with a

single pass of the loop with the patient under local anesthesia in an outpatient setting. Therefore, in most centers, LEEP excision has become the preferred treatment, although laser ablation is still quite adequate if the CIN lesion lies within the range of satisfactory colposcopic assessment. Although the use of cryotherapy in HIV-positive women may seem particularly attractive because of the absence of bleeding and reduced theoretical risk of iatrogenic transmission of HIV, studies have indicated that the use of this modality in HIV-infected women may offer a specific treatment disadvantage. In one study *(30)*, 48% of HIV-infected women with low-grade CIN developed recurrence compared with 1% of seronegative women. Diagnostic procedures, such as cold knife cone conization, diagnostic LEEP, or laser cone, must be performed for the usual indications. Diagnostic procedures may then be considered therapeutic if histologic results are favorable. An algorithm for the management of CIN in HIV-positive patients is presented in Fig. 3.

Recurrence rates for CIN in HIV-positive women with standard therapies have been reported to be as high as 40% at one year and 60% with longer follow-up *(30,31)*. Maiman et al. *(30)* and Del Priore and Lurain *(9)* reported recurrence rates of 39% and 40%, respectively, compared with 9% and 10% on seropositive controls. Fruchter et al. *(32)*, at 36 months, reported a 62% failure rate compared with 18% in control subjects, with an 87% failure rate in patients with CD4 less than 200/mm^3. In addition, progression to higher grade dysplasia was more common in HIV-positive patients as well as multiple episodes of recurrent disease requiring many repeated procedures. It has been well documented that the frequency of recurrence is closely related to immune function. Patients with CD4 counts less than 500/mm^3 are at extremely high risk for recurrence, whereas women with counts greater than 500/mm^3 may be expected to have a recurrence at only twice the rate of seronegative women *(30)*. Therefore, diagnostic and therapeutic strategies may be stratified based on the degree of immunosuppression. Unfortunately, complications after treatment for CIN in HIV-positive women may be increased since one study demonstrated higher rates of excessive bleeding and cervicovaginal infections *(33)*.

While poor treatment results of standard ablative and excisional therapies in HIV-infected women with CIN certainly warrant unique therapeutic strategies, one must recognize that close and meticulous post-therapy surveillance and repetitive aggressive retreatment for persistent and recurrent disease have been successful in preventing progressive neoplasia and invasive cervical carcinoma and should therefore not be abandoned. Hysterectomy, which is rarely used today in the management of CIN in immunocompetent patients, may be considered in individual cases in HIV-seropositive women and be particularly reserved for the multiparous patient with relatively good immune function who has undergone multiple therapeutic procedures for recurrent high dysplasia in whom repeated evaluation is exceedingly difficult.

In response to the high rates of treatment failures for preinvasive cervical disease in HIV-positive women, the AIDS Clinical Trials Group (ACTG) is investigating novel therapeutic approaches. For patients with high-grade dysplasia (CIN-II–III), ACTG examined the role of topical vaginal flourouracil cream maintenance therapy as prophylaxis against recurrent CIN. In this study, patients were randomized to receive standard ablative or excisional therapy alone versus standard therapy plus 2 g of vaginal fluorouracil every 2 weeks for 6 months. Fluorouracil has previously been used with considerable success in the care of immunocompromised women with lower genital tract neoplasia, especially after conventional therapy has resulted in repetitive recurrence. Krebs *(34)* found that prophylactic maintenance therapy with vaginal fluorouracil was effective in the treatment of HPV-associated lesions of the vulva and vagina, especially in immunosuppressed women with multiple organ involvement. Effective local adjunctive therapy with topical chemotherapeutic agents is particularly attractive in HIV-positive patients in light of their favorable therapeutic index, and results of such randomized controlled studies are extremely important.

ACTG 293 is a study involving the treatment of HIV-positive patients with CIN-I. While the standard of care today in immunocompetent women with CIN-I (mild dysplasia) is close follow-up without surgical therapy, it is unknown whether such conservative strategy is safe in seropositive patients in light of reports suggesting more aggressive disease. In this trial, patients are randomized between observation only without any surgical

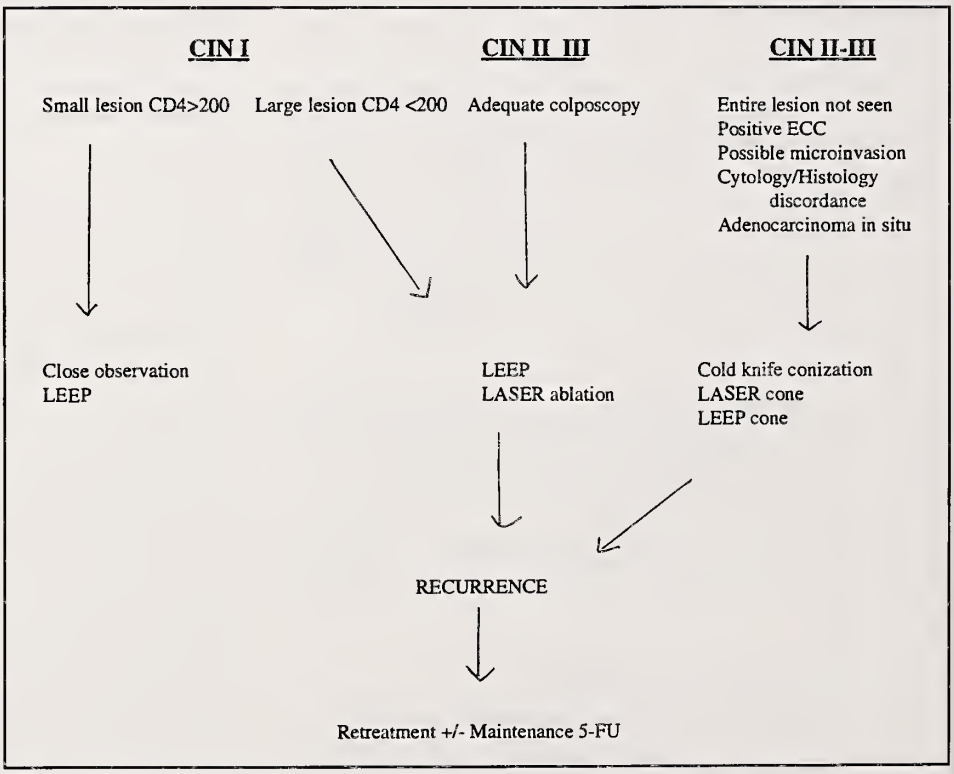


Fig. 3. Treatment options for cervical intraepithelial neoplasia in human immunodeficiency virus-positive women. 5-FU = fluorouracil.

therapy versus oral isotretenoin at a dose of 0.5 mg/kg per day for 6 months, with close attention to follow-up for regression, persistence, or progression. The results of this trial are likely to have implications for women with CIN in general.

Optimization of immune function and lowering of viral load with proven and newer anti-HIV drugs are also desirable, and many patients today are placed on multiple drug regimens. It is unknown whether any of such interventions has any impact on cervical disease, and the presence of multiple confounding variables will make this issue exceedingly difficult to study. At present, the principles of aggressive initial evaluation, more frequent cytologic screening, and meticulous post-therapy surveillance with liberal repeat colposcopy and re-treatment for recurrent dysplasia in managing preinvasive disease in HIV-infected women seem prudent as we investigate novel treatment strategies in clinical trials.

References

(1) Provencher D, Valme, S, Averette HE, Ganjei P, Donato D, Penalver M, et al. HIV status and positive Papanicolaou screening: identification of a high-risk population. Gynecol Oncol 1988;31:184–90.

(2) Schrager LK, Friedland GH, Maude D. Cervical and vaginal squamous cell abnormalities in women infected with the human immunodeficiency virus. J AIDS 1990;2:570–5.

(3) Marte C, Kelly P, Cohen M, Fruchter RG, Sedlis A, Gallo L, et al. Papanicolaou smear abnormalities in ambulatory care sites for women infected with the human immunodeficiency virus. Am J Obstet Gynecol 1992;166: 1232–7.

(4) Feingold AR, Vermund SH, Kelley KF, Schreiber LK, Munk G, Friedland GH, et al. Cervical cytological abnormalities and papillomavirus in women infected with HIV. J AIDS 1990;3:896–903.

(5) Smith JR, Kitchen VS, Bothcherby M, Hepburn M, Wells C, Gor D, et al. Is HIV infection associated with an increase in the prevalence of cervical neoplasia? Br J Obstet Gynaecol 1993;100:149–53.

(6) Maiman M, Tarricone N, Viera J. Suarez J, Serur E, Boyce JG. Colposcopic evaluation of human immunodeficiency virus-seropositive women. Obstet Gynecol 1991;78:84–8.

(7) Kell PD, Shah PM, Barton SE. Colposcopic screening of HIV seropositive women—help or hinderance? Int Conf AIDS 1992;8:B96.

(8) Fink MJ, Fruchter RG, Maiman M, Kelly P, Sedlis A, Weber CA, et al. The adequacy of cytology and colposcopy in diagnosing cervical neoplasia in HIV-seropositive women. Gynecol Oncol 1994;55:133–7.

(9) Del Priore G, Lurain J. The value of cervical cytology in HIV-positive women. Gynecol Oncol 1995;56:395–8.

(10) Korn A, Autry M, DeRemer P. Sensitivity of the Papanicolaou smear in human immunodeficiency virus-infected women. Obstet Gynecol 1994;83: 401–4.

(11) Fruchter RG, Maiman M, Silman FH, Camilien L, Webber CA, Kim DS. Characteristics of cervical intraepithelial neoplasia in women infected with the human immunodeficiency virus. Am J Obstet Gynecol 1994;171: 531–7.

(12) Maiman M, Fruchter RG, Sedlis A, Feldman J, Chen P, Burk RD, et al. Prevalence, risk factors and accuracy of cytologic screening for cervical intraepithelial neoplasia in women with the human immunodeficiency virus. Accepted for publication Am J Obstet Gynecol.

(13) Alloub MM, Barr BB, McLaren KM, Smith GE, Bunney MH, Smart GE. Human papillomavirus infection and cervical intraepithelial neoplasia in women with renal allografts. Br Med J 1989;298:153–6.

(14) Adachi A, Fleming I, Burk RD, Ho GY, Klein RS. Women with human immunodeficiency virus infection and abnormal Papanicolaou smears: a prospective study of colposcopy and clinical outcome. Obstet Gynecol 1993;81:372–7.

(15) Maiman M, Fruchter RG, Guy L, Cuthill S, Levine P, Serur E. Human immunodeficiency virus infection and invasive carcinoma. Cancer 1993; 71:402–6.

(16) Klevens MR, Fleming PL, Mays MA, Frey R. Characteristics of women with AIDS and invasive cancer. Obstet Gynecol 1996;88:169–73.

(17) Maiman M, Fruchter MG, Clark M, Arrastia CD, Matthews R, Gates EJ. Cervical cancer as an AIDS-defining illness. Obstet Gynecol 1997;89: 76–80.

(18) Maiman M, Fruchter RG, Serur E, Remy JC, Feuer G, Boyce JG. Human immunodeficiency virus and cervical neoplasia. Gynecol Oncol 1990;38: 377–82.

(19) Singh GS, Aikins JK, Deger R, King S, Mikuta JJ. Case report on metastatic cervical cancer and pelvic inflammatory disease in the AIDS patient. Gynecol Oncol 1994;54:372–6.

(20) Schwartz LB, Carcangiu ML, Bradham L, Schwartz PE. Rapidly progressive squamous cell carcinoma of the cervix coexisting with human immunodeficiency virus infection: clinical opinion. Gynecol Oncol 1991;41: 255–8.

(21) Relliman MA, Dooley DP, Burke TW, Berkland ME, Longfield RN. Rapidly progressing cervical cancer in a patient with human immunodeficiency virus infection. Gynecol Oncol 1990;36:435–8.

(22) Conti M. Prevalence and risk of progression of genital intraepithelial neoplasia in women with human immunodeficiency virus infection. Adv Gynecol Obstet Res 1991;3:283–7.

(23) Petry KU, Scheffel D, Bode U. Cellular immunodeficiency enhances the progression of human papillomavirus-associated cervical lesions. Int J Cancer 1994;57:836–40.

(24) Sun XW, Ellerbrock TV, Lungu O, Chiasson MA, Bush TJ, Wright TC. Human papillomavirus infection in human immunodeficiency virus-seropositive women. Obstet Gynecol 1995;85:680–6.

(25) Laga M, Icenogle JP, Marsella R, Manoka AT,N Nzila, RW Ryder. Genital papillomavirus infection and cervical dysplasia-opportunistic complications of HIV infection. Int J Cancer 1992;50:45–8.

(26) Johnson J, Burnett A, Willet G, Young MA, Doniger J. High frequency of latent and clinical human papillomavirus cervical infections in immunocompromised human immunodeficiency virus-infected women. Obstet Gynecol 1992;9:321–7.

(27) Schafer A, Friedmann W, Mielke M, Schwartlander B, Koch MA. The increased frequency of cervical dysplasia-neoplasia in women infected with the human immunodeficiency virus is related to the degree of immunosuppression. Am J Obstet Gynecol 1991;164:593–9.

(28) Wright TC, Ellerbrock TV, Chiasson MA, Van Devanter N, Sun XW, and the New York Cervical Disease Study. Cervical intraepithelial neoplasia in women infected with the human immunodeficiency virus: prevalence, risk factors and validity of Papanicolaou smears. Obstet Gynecol 1994;84: 591–7.

(29) Del Priore G, Gilmore PR, Maag T, Warshal DP, Cheon TH. Colposcopic biopsies versus loop electrosurgical excision procedure cone histology in human immunodeficiency virus-positive women. J Reprod Med 1996;41: 653–7.

(30) Maiman M, Fruchter RG, Serur E, Levine PL, Arrastia C, Sedlis A. Recurrent cervical intraepithelial neoplasia in human immunodeficiency virus-seropositive women. Obstet Gynecol 1993;82:170–4.

(31) Wright T, Koulos J, Schnoll F, Swanback J, Ellerbrock TV, Chiasson MA, et al. Cervical intraepithelial neoplasia in women infected with the human immunodeficiency virus: outcome after loop electrosurgical excision. Gynecol Oncol 1994;55:253–8.

(32) Fruchter RG, Maiman M, Sedlis A, Bartley L, Camilien L, Arrastia CD. Multiple recurrences of cervical intraepithelial neoplasia in women with human immunodeficiency virus. Obstet Gynecol 1996;87:338–44.

(33) Cuthili S, Maiman M, Fruchter RG, Lopatinsky I, Cheung CC. Complications after treatment of cervical intraepithelial neoplasia in women infected with the human immunodeficiency virus. Reprod Med 1995;40: 823–8.

(34) Krebs HB. Prophylactic topical 5-fluorouracil following treatment of human papillomavirus-associated lesions of the vulva and vagina. Obstet Gynecol 1986;68:837.

Human Herpesvirus Type 8 and Kaposi's Sarcoma

*Robin A. Weiss, Denise Whitby, Simon Talbot, Paul Kellam, Chris Boshoff**

Kaposi's sarcoma-associated herpesvirus or human herpesvirus type 8 (HHV-8) is present in all forms of Kaposi's sarcoma (KS) as well as in primary effusion lymphomas and some cases of Castleman's disease. In KS tissues, HHV-8 is present in endothelial and spindle cells. Current serologic tests suggest that HHV-8 is predominantly found in those at risk of KS and is not as widespread as most other human herpesviruses. HHV-8 encodes various proteins that may play a role in promotion of cellular growth, including cyclin- and G-coupled protein receptor homologues, and anti-apoptotic proteins, including Bcl-2, IL-6 (i.e., interleukin 6), and FLIP (i.e., FLICE inhibitory protein) homologues. In addition, HHV-8 encodes two macrophage inflammatory-like proteins with anti-human immunodeficiency virus and angiogenic potential. [Monogr Natl Cancer Inst 1998;23: 51–54]

This conference is a timely opportunity to focus on the cancers that occur more frequently in acquired immunodeficiency syndrome (AIDS). Kaposi's sarcoma (KS) with *Pneumocystis carinii* infection heralded the AIDS epidemic in its earliest days. KS is seen frequently in Africa and is linked both to human immunodeficiency virus (HIV) (the epidemic form) and HIV-negative KS (the endemic form). It is also endemic in northern Mediterranean and eastern European countries. We shall focus on human herpesvirus type 8 (HHV-8)/Kaposi's sarcoma-associated herpesvirus (KSHV), but many more details are presented later in the conference when molecular biologists analyze the viral genome and describe how it replicates and functions.

The study of HHV-8/KSHV started less than 3 years ago when Yuan Chang and colleagues announced its discovery in December 1994 *(1)*. Since then, about 200 publications on this new herpesvirus have appeared, which indicates the growing interest in it. Without this discovery through the representational difference analysis, this conference might not have been convened. Chang et al. *(1)* called the new virus "Kaposi's sarcoma-associated herpesvirus (KSHV)" because they discovered it in KS tissue of patients with AIDS. Schulz and Weiss *(2)*, in a commentary on their discovery, suggested that "HHV-8" was a slightly more neutral name regarding etiology. The name does not really matter and, throughout the conference, some participants use "KSHV" and others say "HHV-8." It has been argued that, because this virus is associated not only with KS but also with other lesions, such as lymphoma, we should not call it "KSHV." But on that account we should not call HTLV-1 "human T-cell leukemia virus" anymore because it also causes the neurologic disease tropical spastic paraparesis; we should rename "hepatitis G virus" because it does not appear to cause hepatitis.

Table 1. Detection of human herpesvirus type 8 in Kaposi's sarcoma (KS) tissue by polymerase chain reaction*

	Positive	Tested	%
Acquired immunodeficiency syndrome-associated KS	108	113	96
Classic KS	46	49	94
Posttransplant KS	9	9	100
Human immunodeficiency virus-negative gay men	5	5	100
African endemic	41	47	87

*Data compiled from references *(3,29–32)*.

After Chang et al. *(1)* first identified HHV-8 in lesions of KS of AIDS patients, we examined the presence of HHV-8 among the different KS epidemiologic groups. Data on the polymerase chain reaction (PCR) detection of HHV-8 in paraffin-imbedded KS tissue are shown in Table 1. The great majority of KS samples were positive for the presence of this virus, whatever the epidemiologic group *(3,4)*. If we exclude the ones that had poor-quality DNA, as judged by weak amplification of a cellular gene, then virtually 100% of KS samples were HHV-8 positive. Several other groups have accrued similar results. Thus, there is genuine consensus that this virus is universally present in KS lesions.

When we looked for HHV-8 by *in situ* techniques, we observed that a majority of spindle cells showed the presence of HHV-8, although negative nuclei were also present *(5)*. Again, several other groups have confirmed this result. For example, the presence of HHV-8 genes, antigens, and indeed viruses particles has been demonstrated by Orenstein et al. *(6)*.

A continuing question is whether HHV-8 plays a causal role or whether it is a passenger virus. Another question is whether HHV-8, like Epstein-Barr virus, is ubiquitous in the human population or is restricted to KS groups? We favor a causal role, and a consensus is emerging that HHV-8 contributes to the pathogenesis of KS.

**Affiliation of authors:* Institute of Cancer Research, Chester Beatty Laboratories, London, U.K.

Correspondence to: Robin A. Weiss, Ph.D., Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Rd., London SW3 6JB, U.K.

See "Notes" following "References."

© Oxford University Press

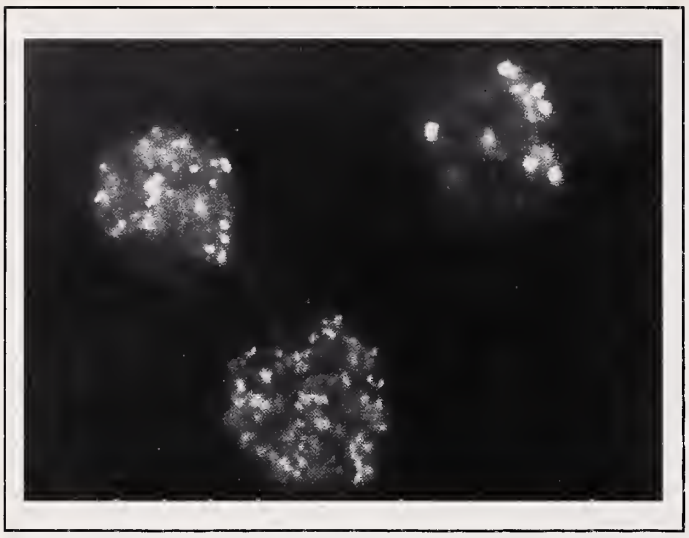


Fig. 1. Nuclear stippling seen in human herpesvirus type 8 latently infected cell line (BCP-1) *(12)* with sera of patients with Kaposi's sarcoma.

Using PCR techniques, some groups reported that the virus is more widespread and detected KSHV in non-KS skin tumors in immunosuppressed patients and in the semen of healthy donors. We could not detect HHV-8 in squamous cell carcinomas of the skin in immunosuppressed transplant recipients *(7)* or in the semen of healthy donors *(8)*.

However, one could argue that this virus is ubiquitous and that it might become activated by angiogenesis but that it is normally latent at very low load or not expressed at all until it is activated by angiogenic cytokines or chemokines. We could not detect HHV-8 in benign hemangiosarcomas or granulomas; as a result, among angiogenic lesions *(3)*, its expression seems to be specific to KS, although the cytokine profiles in these lesions might differ.

Nevertheless, HHV-8 is associated with other lymphoproliferative lesions. Multicentric Castleman's disease was first reported by Soulier et al. *(9)* in France to contain HHV-8, with 14 of 14 HIV-positive patients with Castleman's disease and a rather smaller proportion of HIV-negative patients being HHV-8 positive; we have obtained similar results, as have Dupin et al. *(10)*, among others. Multicentric Castleman's disease is a lymphoproliferative lesion that frequently includes angiogenic proliferation and sometimes presents together with KS itself.

Better known are the primary effusion lymphomas (or body cavity-based lymphomas) first described as containing HHV-8 by Cesarman et al. *(11)*. These lymphomas tend to grow as effusions in the pleural or pericardial cavity. They lack many of the cell surface adhesion molecules that one might expect to be

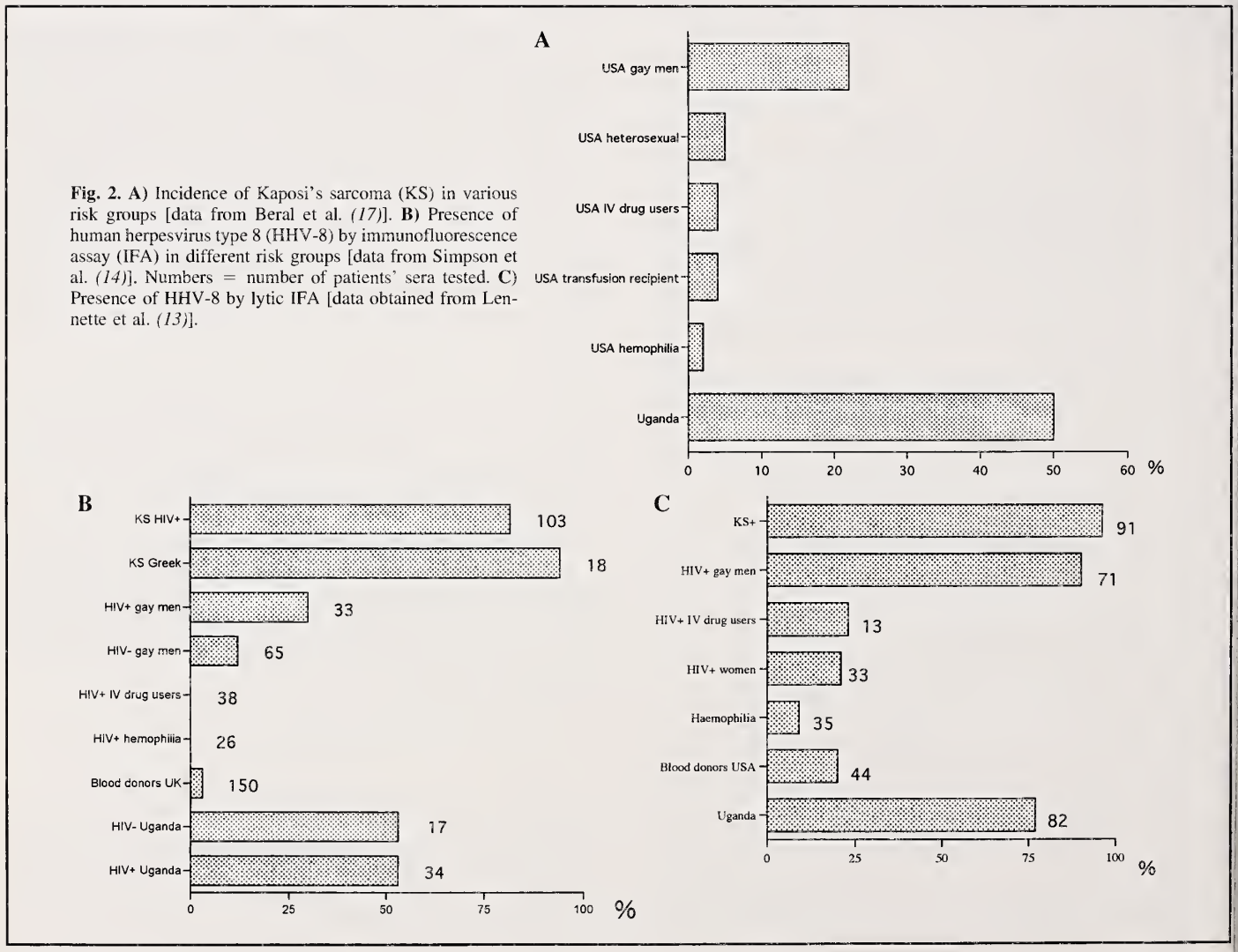


Fig. 2. A) Incidence of Kaposi's sarcoma (KS) in various risk groups [data from Beral et al. *(17)*]. **B)** Presence of human herpesvirus type 8 (HHV-8) by immunofluorescence assay (IFA) in different risk groups [data from Simpson et al. *(14)*]. Numbers = number of patients' sera tested. **C)** Presence of HHV-8 by lytic IFA [data obtained from Lennette et al. *(13)*].

present on lymphomas *(12)*. While many primary effusion lymphomas are positive for EBV, practically all of them are positive for HHV-8. Primary effusion lymphoma cell lines have provided the basis for many studies presented at this conference on the induction, replication, and cloning of HHV-8 genes, as well as for HHV-8 serology.

With the use of primary effusion lymphoma cell lines in culture with latent infection as an assay for immunofluorescence antibody detection, sera from patients with KS show a punctate nuclear staining (Fig. 1). This assay has been the basis of several studies for first-generation serologic surveys of HIV infection *(13–16)*.

These cell lines also can be induced by phorbol esters or sodium butyrate to enter a lytic viral replication in which many more of the HHV-8 proteins are expressed, and hence more antigens are there to be recognized. This induction of lytic viral protein expression should produce greater sensitivity in detection of antibodies, but the question is whether it reduces specificity. Recombinant antigen assays are not yet as sensitive as immunoflurescence assays *(14)*.

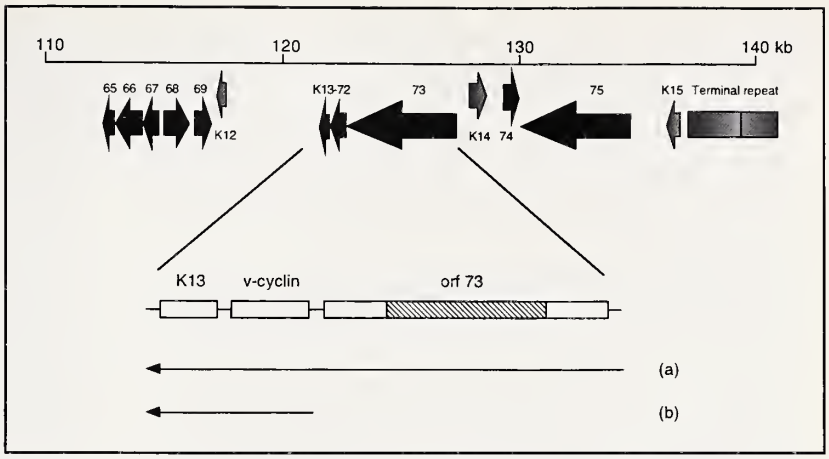


Fig. 3. Human herpesvirus type 8 LNA-1 (latent nuclear antigen) is encoded by orf 73 and forms part of a larger messenger RNA (mRNA) transcript including orf 72 (cyclin) and K13 (v-FLIP). During latent infection, orf 73 (LNA-1) is transcribed with the viral cyclin and anti-apoptotic protein K13 (a). During lytic induction, orf 73 is spliced out and K13 and cyclin (b) are transcribed as a bicistronic mRNA [data from Kellam et al. *(22)*].

As discussed at the beginning of this conference, the incidence of KS is different in different parts of the world and in different categories of patients with AIDS *(17)*. Our serology for HHV-8 antibodies fits the same pattern (Fig. 2). Among all the different categories of AIDS patients from the United States and Northern Europe, we can see that HHV-8 infection occurs most frequently in gay men and least often in patients with hemophilia. Heterosexual acquisition of HIV in Africa is also associated with KS.

We showed in 1995 that HHV-8 genomes are detectable in the blood of a proportion of KS-negative, HIV-positive gay men and that their presence is strongly predictive of later development of KS *(18)*. If we consider serology and genome detection together, there appears to be a strong link between HHV-8 infection and later development of KS. Geographically, too, the detection of HHV-8 antibodies is associated with high-incidence areas for KS (e.g., Greece, Italy, and Uganda) *(19)*. In a collaboration with Freddy Sits in Johannesburg, South Africa, we have shown that mother-to-child transmission is an important route for acquiring HHV-8 in Africa *(20)*.

The antigen specifying the nuclear speckle staining used in immunofluorescence studies has recently been identified as the HHV-8 orf 73 product *(21,22)*. This gene has a complex pattern of transcription, differing in latent and lytic infections. In latent infection, a large transcript including v-cyclin (orf 72) and K13 FLIP (i.e., FLICE inhibitory protein) (orf 71) is evident (Fig. 3) *(22)*.

With regard to v-cyclin, we together with Mittnacht, Chang, and Moore *(23,24)*, have shown that the v-cyclin protein can phosphorylate retinoblastoma protein. This phosphorylation is mediated mainly through cdk6 *(24,25)*. HHV-8 v-cyclin can also block the function of the Rb protein as a tumor suppressor in a transfection assay in osteosarcoma cells that lack Rb *(23)*. This is an example of one HHV-8 gene that could be potentially oncogenic. Others include the genes for a G-coupled protein receptor, an interferon response element *(26,27)*, and apoptosis inhibitors such as the v-FLIP (K13), v-bcl-2, and v-IL-6. HHV-8 also encodes for chemokine homologues (v-MIP-I and v-MIP-II), which we have shown to have HIV-inhibitory activity *(26)* and angiogenic potential in an *in vivo* assay *(28)*.

Several of these genes are discussed in more detail in other papers delivered at this conference. A caveat is that the presence of transforming genes in a human virus does not necessarily mean that the virus is oncogenic in humans. The adenovirus genes EIA and EIB are classical oncogenes, but there is no evidence that adenoviruses cause tumors in humans. The same is true for BK papovavirus. Thus, we have to be cautious about interpreting our findings with genes when we clone them and express them in NIH3T3 cells or other experimental systems. What makes HHV-8 look increasingly guilty of causing KS and lymphoma is the combination of epidemiologic and experimental data: The prevalence of the virus is generally high in those human populations with a high incidence of KS, and the virus carries the tools to stimulate cell proliferation and to induce neovascularization.

References

(1) Chang Y, Cesarman E, Pessin MS, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 1994;266:1865–9.

(2) Schulz TF, Weiss RA. Kaposi's sarcoma. A finger on the culprit. Nature 1995;373:17–8.

(3) Boshoff C, Whitby D, Hatziioannou T, et al. Kaposi's-sarcoma-associated herpesvirus in HIV-negative Kaposi's sarcoma. Lancet 1995;345:1043–4.

(4) Chang Y, Ziegler J, Wabinga H, et al. Kaposi's sarcoma-associated herpesvirus and Kaposi's sarcoma in Africa. Arch Intern Med 1996;156:202–4.

(5) Boshoff C, Schulz TF, Kennedy MM, et al. Kaposi's sarcoma-associated herpesvirus infects endothelial and spindle cells. Nat Med 1995;1:1274–8.

(6) Orenstein JM, Alkan S, Blauve HA, et al. Visualization of human herpesvirus-8 in Kaposi's sarcoma by light and transmission electron microscopy. AIDS 1997;11:735–45.

(7) Boshoff C, Talbot S, Kennedy M, et al. HHV8 and skin cancers in immunosuppressed patients. Lancet 1996;347:338–9.

(8) Howard MR, Whitby D, Bahadur G, et al. Detection of human herpesvirus 8 DNA in semen from HIV-infected individuals but not healthy semen donors. AIDS 1997;11:F15-F19.

(9) Soulier J, Grollet L, Oksenhendler E, et al. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. Blood 1995;86:1276–80.

(10) Dupin N, Gorin I, Deleuze J, et al. Herpes-like DNA sequences, AIDS related tumors, and Castleman's disease. N Engl J Med 1995;333:798–9.
(11) Cesarman E, Chang Y, Moore PS, et al. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. N Engl J Med 1995;332:1186–91.
(12) Boshoff C, Gao SJ, Healy LE, et al. Establishing a KSHV positive cell line (BCP-1) from peripheral blood and characterizing its growth in Nod/SCID mice. Blood. In press.
(13) Lennette ET, Blackbourn DJ, Levy JA. Antibodies to human herpesvirus type 8 in the general population and in Kaposi's sarcoma patients. Lancet 1996;348:858–61.
(14) Simpson GR, Schulz TF, Whitby D, et al. Prevalence of Kaposi's sarcoma associated herpesvirus infection measured by antibodies to recombinant capsid protein and latent immunofluorescence antigen. Lancet 1996;348: 1133–8.
(15) Kedes DH, Operskalski E, Busch M, et al. The seroepidemiology of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus): distribution of infection in KS risk groups and evidence for sexual transmission. Nat Med 1996;2:918–24.
(16) Gao SJ, Kingsley L, Li M, et al. KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. Nat Med 1996; 2:925–8.
(17) Beral V, Peterman TA, Berkelman RL, et al. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? Lancet 1990;335: 123–8.
(18) Whitby D, Howard MR, Tenant Flowers M, et al. Detection of Kaposi sarcoma associated herpesvirus in peripheral blood of HIV-infected individuals and progression to Kaposi's sarcoma. Lancet 1995;346:799–802.
(19) Whitby D, Luppi M, Barozzi P, et al. Human herpesvirus 8 seroprevalence in blood donors and lymphoma patients from different regions of Italy. J Natl Cancer Inst 1998;90:395–397.
(20) Bourboulia D, Whitby D, Boshoff C, et al. Serological evidence for vertical transmission of KSHV/HHV-8 in healthy South African children. Manuscript submitted for publication.
(21) Rainbow L, Platt GM, Simpson GR, et al. The 222–234 kDa nuclear protein (LNA) of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) is encoded by orf 73 and a component of the latency-associated nuclear antigen. J Virol 1997;71:5915–21.
(22) Kellam P, Boshoff C, Whitby D, et al. Identification of a major latent nuclear antigen (LNA-1) in the human herpesvirus 8 (HHV-8) genome. J Hum Virol 1997;1:19–29.
(23) Chang Y, Moore PS, Talbot SJ, et al. Cyclin encoded by KS herpesvirus. Nature 1996;382:410.
(24) Godden-Kent D, Talbot SJ, Boshoff C, et al. The cyclin encoded by Kaposi's sarcoma-associated herpesvirus (KSHV) stimulates cdk6 to phosphorylate the retinoblastoma protein and histone H1. J Virol 1997;71: 4193–8.
(25) Li M, Lee H, Yoon DW, et al. Kaposi's sarcoma-associated herpesvirus encodes a functional cyclin. J Virol 1997;71:1984–91.
(26) Moore PS, Boshoff C, Weiss RA, et al. Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV. Science 1996; 274:1739–44.
(27) Gao SJ, Boshoff C, Jayachandra S, et al. KSHV ORF K9 (vIRF) is an oncogene which inhibits the interferon signaling pathway. Oncogene 1997; 15:1979–85.
(28) Boshoff C, Endo Y, Collins PD, et al. Angiogenic and HIV inhibitory functions of KSHV-encoded chemokines. Science 1997;278:290–4.
(29) Chang Y, Ziegler J, Wabinga H, et al. Kaposi's sarcoma-associated herpesvirus and Kaposi's sarcoma in Africa. Uganda Kaposi's Sarcoma Study Group. Arch Intern Med 1996;156:202–4.
(30) Schalling M, Ekman M, Kaaya BE, et al. A role for a new herpes virus (KSHV) in different forms of Kaposi's sarcoma. Nat Med 1995;1:705–6.
(31) Dupin N, Grandadam M, Calvez V. Herpes-like DNA sequences in patients with Mediterranean Kaposi's sarcoma. Lancet 1995;345:761–2.
(32) Moore PS, Chang Y. Detection of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma lesions from persons with and without HIV infection. N Engl J Med 1995;332:1181–5.

Notes

Supported by the Cancer Research Campaign and the Medical Research Council.

The work presented here is part of a much larger consortium of collaborators in epidemiology and molecular pathology as seen in the authorship of our primary publications.

Some Aspects of the Pathogenesis of HIV-1-Associated Kaposi's Sarcoma

*Robert C. Gallo**

Kaposi's sarcoma (KS) is a very rare tumor except after human immunodeficiency virus type 1 (HIV-1) infection when it becomes common. Most investigators assume that the role of HIV-1 is passive, i.e., via inducing immune deficiency, thereby enhancing cancer development, and specifically, in the case of human herpesvirus 8 (KSHV) by enhancing HHV-8 replication. We suggest that the role of HIV-1 is more active in the disease process by at least two events: 1) promoting an increase in inflammatory cytokines, which through sustained release influences early stage KS by inciting local microinflammatory responses, and 2) by the Tat protein that effects growth of the inflammatory cells. Cultures of all activated endothelial spindle cells, whether hyperplastic or neoplastic, are negative for HHV-8; transmission of HHV-8 does not induce cell growth or transformation; monkeys immune suppressed by simian immunodeficiency virus infection and infected also with HHV-8 do not develop KS; polymerase chain reaction analysis of blood cells shows HHV-8 sequences in monocytes and B cells (about 20% of normal donors in Maryland); M. Starzl showed that early KS has few cells (mostly macrophage) positive for HHV-8, increasing and present in endothelial cells only late in the disease; no increase in HHV-8 has been found in association with progressive immune deficiency; and recent studies in Gambia by others showed that HHV-8 is a very common infection, and though HHV-2 is known to be relatively common, HIV-1 is unusual and so is KS. Collectively, these findings lead me to conclude that there is little evidence that HHV-8 is a transforming virus as has been repeatedly suggested and that its role in KS is more likely to be indirect (like HIV-1), perhaps necessary but hardly likely to be sufficient for KS development. [Monogr Natl Cancer Inst 1998;23:55–57]

Beginning in 1987, my colleagues and I have studied Kaposi's sarcoma (KS) cells in culture and after xenotransplantation to nude mice in an effort to gain insight into the nature of KS tumor cells and possibly to shed light on its pathogenesis *(1,2)*. Our findings are consistent with the idea that the KS spindle cells (SCs) are chiefly activated vascular endothelial cells (ECs). The vast majority of successfully cultured cells were normal diploid, grew only transiently in nude mice, and induced angiogenesis in these animals by release of cytokines *(1,2)*. We noted that inflammatory cytokines—interleukin 1 (IL-1), IL-6, oncostatin M, and interferon gamma (IFN γ), known to be elevated in HIV-1-infected persons *(3)*—promoted growth of these cells *(1,4)*. Interestingly, IFN γ can convert ECs from cuboidal to spindle shape and promotes the migration of these cells into the circulation, presumably in search of an inflammatory lesion. IFN γ also activates the expression of integrins on EC as well as several cytokines *(5)*. The "culturable" KS SCs themselves produce high levels of cytokines, such as platelet-derived growth factor (PDGF), IL-6, IL-1, and, notably, basic fibroblast growth factor (bFGF) and vascular endothelial cell growth factor (VEGF), both of which are potent angiogenic molecules *(6,7)*. We have provided data that bFGF is important to sustain growth of the hyperplastic cells *(2,4)*, and Masood and colleagues *(8)* provided evidence that VEGF is important for neoplastic cells. The inflammatory cytokines also promote increased expression of adhesion molecules that facilitate greater interactions of leukocytes and ECs. This includes HIV-1-infected CD4 T cells and macrophages. We showed that the Tat protein of HIV-1 is actively released into the extracellular fluid *(9)*, and others showed that it is taken up by nearby cells *(9–11)*. We showed it is also taken up by KS SCs *(12)*. Our results indicate that Tat, at levels released by cells, promotes the migration, invasiveness, and growth of KS SC and that these effects are mediated by the RGD motif of Tat and by its basic region *(11,12)*; the RGD region, through its molecular mimicry of fibronectin and vitronectin, interacts with integrins, whereas the basic region, by competing with bFGF for binding to heparin sulfate proteoglycans, allows more free soluble bFGF available for growth promotion *(13)*. ECs require two signals for growth: adhesion and growth promotion. Both are augmented by HIV-1 infection.

Whether KS is hyperplastic or neoplastic remains debatable. There are results and arguments for both. Reports indicate that many KS patients have tumors which are clonal *(14)*, and we and others have succeeded in isolating neoplastic clones from a few patients (KS Y-1 and KS SLK, respectively) *(15,16)*. Of possible importance is the result showing similar chromosomal markers in both neoplastic cell lines *(17)*. At this time, we can conclude that a significant component of cells of a KS lesion (and perhaps all cells in some patients) are hyperplastic, whereas neoplastic cells can be found in some. One interpretation is that all KS lesions have neoplastic cells from the start, but akin to the Reed-Sternberg cells of Hodgkin's disease they recruit normal cells and remain in the minority but in KS are not morphologically distinguishable as are the Reed-Sternberg cells. Alternatively, KS may begin as a hyperplasia, and only in some patients does a neoplastie transformation occur. We may be able to obtain a definitive answer to this problem if we are able to generate a specific probe based on the karyotypic common abnormality

**Correspondence to:* Robert C. Gallo, M.D., Institute of Human Virology, University of Maryland at Baltimore, 725 W. Lombard St., 3rd Floor, Baltimore, MD 21201-1192.

© Oxford University Press

found in the short arm of chromosome 3 in the two neoplastic cell lines *(17)*.

KS is rare and usually not aggressive except in association with HIV-1 infection. Most investigators suggest that the role of HIV-1 is passive, i.e., in inducing immune suppression. In some unexplained way, this is supposed to increase the incidence of cancer. However, with the exception of lymphomas, this has not been demonstrated. Others, such as R. Weiss (these meetings and report from the October 1995 Pasteur Vaccine meeting), suggested that HIV-1-induced immune suppression may favor HHV-8 replication. However, there is no evidence for increased HHV-8 with increasing immune suppression. Furthermore, it is clear that immune suppression occurs without KS development, e.g., congenital, chemotherapy-induced, radiation-therapy induced, or, more to the point, after HIV-2 infection or SIV in monkeys, both of which may produce immune suppression but neither promoting KS. It is of interest, in this regard, that the Tat of SIV and HIV-2 lacks the RGD domain. Finally, KS can occur in the absence of any known immune suppression; e.g., it can be the first sign of HIV-1 infection. Additionally, in non-HIV-1, KS-associated immune suppression is generally not found, e.g., African endemic KS. It is also unlikely that it is a significant feature of classical KS. As noted above, we suggest that the role of HIV-1 in KS is active in its promotion of inflammatory cytokines and by its Tat protein *(3)*, thereby creating microvascular inflammatory lesions that I suggest are the pre-KS lesions.

HHV-8 (KSHV) and KS

This virus must be involved in the cause of KS if it is not ubiquitous (which seems to be the case), because it is almost always found in KS lesions. Before a final acceptance of this, however, more refined epidemiologic data with HHV-8 are needed. Originally, it was said by Cesarman et al. *(18)* to be very rare and virtually limited to KS, but soon after, a higher prevalence was found *(19)*, which seems to be steadily edging upward. A true understanding of its prevalence is probably not yet known. I believe a coordinated study by polymerase chain reaction analysis of separated blood cell populations in multiple laboratories of multiple populations will be the best answer to this problem. Currently, we can estimate it as at least 5% of most populations and likely substantially higher. Recent serology of Gambians by R. Weiss (reported at the International Meeting on Herpes 6, 7 and 8 in Pisa, Italy, June 1997) showing HHV-8 infection in at least 64% of people is disturbing because KS is unusual in Gambia. Of interest and compatible with our hypothesis is that, although HIV-2 is common in Gambia, HIV-1 is unusual.

Chang and Moore (these meetings and elsewhere) have argued that HHV-8 is a transforming virus. Mostly, this has been based on indirect suggestions: 1) its presence in KS lesions; 2) its relatedness to Epstein-Barr virus (EBV) and herpes saimiri; and 3) its genome containing homologues of cellular genes, some of which are involved in cell growth. However, EBV very rarely causes cancer and as Fleckenstein noted (these meetings), neither herpes saimiri or any Rhadinovirus (like HHV-8) have ever been shown to cause disease in their own host. Finally, having some genome homology does not make HHV-8 equivalent to these viruses. Needless to emphasize, some homology to cellular genes says nothing about the capacity to transform. It only offers leads to a virus known to be transforming.

We have been impressed by the fact that, when directly studied, HHV-8 has not only not transformed cells, it has not promoted growth of any primary cells and most importantly not of EC. Moore (this meeting) described DNA of HHV-8 transforming NIH 3T3 cells. However, the limitations of this assay are well known, and the data are not yet available for scrutiny. Moreover, this artificial assay cannot be used to say a virus is transforming. The NIH 3T3 cells are already immortal, and many kinds of DNA lead to these cells promoting tumors in nude mice. Indeed, DNA from HIV-6 can also do the same as well as DNA of some adenoviruses, neither of which is implicated in human cancer.

As already noted, our cultured hyperplastic cells (KS #1–14) and all known cultured neoplastic KS cell lines (KS Y-1 and KS SLK) are HHV-8 negative *(20)*. Furthermore, monkeys infected with SIV (immune deficient as a result) and also infected with HHV-8 do not develop KS (unpublished results of Colombini in our group in collaboration with Biberfeld). These findings and the recent Gambia data (low KS, high prevalence of HHV-8) would lead me to the hypothesis that HHV-8 is not likely to cause KS as a transforming virus. If ultimately found to be ubiquitous, it could still be only a passenger virus. However, this seems unlikely. I favor the idea that it plays an enhancing and likely an essential role in the origin of KS, albeit like HIV-1 likely to be an indirect one, i.e., not simply by infecting a cell and transforming this cell into a neoplastic cell. Instead, I suggest that HHV-8 is recruited to microinflammatory lesions and promotes growth of such lesions (by augmentation of cytokine release?). Our preliminary results indicate that the same inflammatory cytokines which increased after HIV-1 infection and which themselves promote microinflammatory changes also promote HHV-8 replication (unpublished observations of S. Colombini). The converse may also be true and needs testing; i.e., does HHV-8 stimulate the production of growth-promoting cytokines bFGF, VEGF, IL-6, etc., and/or promote growth by contributing, through its own gene products, homologs of cellular inflammatory cytokines?

References

(1) Salahuddin SZ, Nakamura S, Biberfeld P, Kaplan MH, Markham PD, Larsson L, Gallo RC. Angiogenic properties of Kaposi's sarcoma-derived cells after long-term culture *in vitro*. Science 1988;242:430–3.

(2) Ensoli B, Nakamura S, Salahuddin SZ, Biberfeld P, Larsson L, Beavor B, et al. AIDS-Kaposi's sarcoma-derived cells express cytokines with autocrine and paracrine growth effects. Science 1989;243:223–6.

(3) Samaniego F, Gallo RC. Immunopathogenesis of Kaposi's sarcoma. In: Gupta E, editor. Immunopathogenesis of HIV infection. New York: Plenum Press, 1996:437–50.

(4) Samaniego F, Markham PD, Gallo RC, Ensoli B. Inflammatory cytokines induce AIDS-KS-derived spindle cells to produce and release bFGF and to enhance KS-like lesion formation in nude mice. J Immunol 1995;154: 3582–92.

(5) Seigneur M, Constans J, Blann A, Renard M, Pellegrin JL, Amiral J, et al. Soluble adhesion molecules in endothelial cell damage in HIV infected patients. Thromb Haemost 1997;77:646–9.

(6) Samaniego F, Markham PD, Gendelman R, Gallo RC, Ensoli B. Vascular endothelial growth factor and basic fibroblast growth factor are expressed in both AIDS-associated and classical Kaposi's sarcoma (KS) and syner-

gize in inducing angiogenesis, vascular permeability and KS lesion development. Induction by inflammatory cytokines. AIDS Res Hum Retroviruses 1995;11:S95.

(7) Samaniego F, Markham PD, Gendelman R, Gallo RC, Ensoli B. Inflammatory cytokines induce endothelial cells to produce and release basic fibroblast growth factor and to promote Kaposi's sarcoma-like lesions in nude mice. J Immunol 1997;158:1887–94.

(8) Masood R, Cai J, Zheng T, Smith DL, Naidu Y, Gill PS. Vascular endothelial growth factor/vascular permeability factor is an autocrine growth factor for AIDS-Kaposi sarcoma. Proc Natl Acad Sci U S A 1997;94:979–84.

(9) Ensoli B, Buonaguro L, Barillari G, Fiorelli V, Gendelman R, Morgan RA, et al. Release, uptake and effects of extracellular human immunodeficiency virus type 1 Tat protein on cell growth and viral transactivation. J Virol 1993;67:277–87.

(10) Frankel AD, Pabo CO. Cellular uptake of the Tat protein from human immunodeficiency virus. Cell 1988;55:1189–93.

(11) Westendrop MO, Frank R, Ochsenbauer C, Stricker K, Dhein J, Walezak H, et al. Sensitization of T cells to CD95-mediated apoptosis by HIV-1 Tat and gp120. Nature 1995;375:497–500.

(12) Ensoli B, Gendelman R, Markham P, Fiorelli V, Colombini S, Raffeld M, et al. Synergy between basic fibroblast growth factor and HIV-1 Tat protein in induction of Kaposi's sarcoma. Nature 1994;371:674–80.

(13) Chang H, Samaniego F, Buonaguro L, Bair BC, Gallo RC, Ensoli B. Extracellular Tat binds heparin-sulfate complexes of the cell surface and extracellular matrix (ECM) through the basic region. AIDS. In press.

(14) Rabkin CS, Janz S, Lash A, Coleman AE, Musuba E, Liotta L, et al. Monoclonal origin of multicentric Kaposi's sarcoma lesions. N Engl J Med 1997;336:988–93.

(15) Lunardi-Iskandar Y, Bryant J, Zeman R, Samaniego F, Larn VH, Thierry A, et al. A pregnancy hormone human chorionic gonadotropin blocks tumorigenesis and metastasis of Kaposi's sarcoma cells (KS Y-1) in immunodeficient mice. Nature 1995;375:64–8.

(16) Herndier BG, Werner A, Arnstein P, Abbey NW, Demartis F, Cohen RL, et al. Characterization of a human Kaposi's sarcoma cell line that induces angiogenic tumors in animals. AIDS 1994;8:575–81.

(17) Popescu NC, Zimonjie DB, Leventon-Kriss S, Bryant JL, Lunardi-Iskander Y, et al. Deletion and translocation involving chromosome 3 (p14) in two tumorigenic Kaposi's sarcoma cell lines. J Natl Cancer Inst 1996;88: 450–5.

(18) Cesarman E, Chang Y, Moore PS, Said JW, Knowles DM. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. N Engl J Med 1995;332:1186–91.

(19) Monini P, deLellis L, Fabris M, Rigolin F, Cassai E. Kaposi's sarcoma-associated herpesvirus DNA sequences in prostate tissue and human semen. N Engl J Med 1996;334:1168–72.

(20) Flamand L, Zeman RA, Bryant JL, Lunardi-Iskandar Y, Gallo RC. Absence of human herpesvirus 8 DNA sequences in neoplastic Kaposi's sarcoma lines. J AIDS 1996;13:194–7.

''inical Overview: Issues in Kaposi's .rcoma Therapeutics

*san E. Krown**

ur questions are posed that are critical to the development .nproved therapeutic and prophylactic strategies for Ka-.'s sarcoma (KS). 1) Can we predict who will develop KS? ccurate identification of high-risk factors for KS develop-ent is essential for the development of KS prophylaxis tri-s. 2) Can developing insights into KS pathogenesis be ·nslated into improved therapeutic and/or new prophylac- strategies for patients at high risk? Several approaches e being developed that target new blood vessel develop-nt, inflammatory cytokines, and the viruses that are im-.cated in KS pathogenesis. 3) How does the improved prog-is for human immunodeficiency virus (HIV)-infected .ents affect KS treatment strategy? Improved anti-HIV rapy has implications for the timing of KS therapy, the .ce of therapeutic approaches, and the potential for ad-se drug interactions. 4) How can we best evaluate benefits KS treatment? More rigorous, standardized criteria in development and will be essential not only for accu- documentation of objective tumor regression, but also assessment of tumor-associated symptom relief in a titative, function-oriented way. [Monogr Natl Cancer 1998;23:59–63]

emendous progress has been made recently in identifying factors that contribute to the development of Kaposi's sar-a (KS), the most common cancer associated with human unodeficiency virus type 1 (HIV-1) infections, but treatment S remains suboptimal. In reviewing the issues in KS man-.nent, the following four critical questions must be addressed attempts are made to improve therapy and develop KS pro-ylaxis strategies over the next several years:

- Can we predict who will develop KS?
- Can developing insights into KS pathogenesis be translated into improved therapy and/or new prophylactic strategies for patients at high risk?
- How do the improved treatments and prognosis for HIV-infected patients affect our overall KS treatment strategy?
- ow can we best evaluate benefits from KS treatment?

These issues will be briefly discussed in the succeeding sec-ons of this article.

rediction of KS Development

The ability to predict who will develop KS is critical for the evelopment and interpretation of prophylaxis trials. The poten-l to intervene prior to the development of KS is an exciting w possibility, but some prophylactic strategies may involve e administration of drugs with toxic potential, so accurate identification of high-risk individuals is essential. Although we have known for a long time that, in developed countries, HIV-infected gay and bisexual men are most likely to develop KS, only a minority do so. In developed countries, however, it is often not appreciated that HIV-infected individuals from other transmission groups also show a marked increase in KS risk that has recently been estimated in the United States as 10 000 times greater than that in the age- and sex-matched general population *(1)*. Therefore, sexual orientation and behavior alone are not sufficient to identify who will develop KS in the United States, and they are not relevant factors in other parts of the world, like Africa, where KS incidence is relatively high in both sexes and is not associated with male homosexuality.

Although human herpesvirus 8 (HHV-8; Kaposi's sarcoma-associated herpesvirus) has not been established as the cause of KS (i.e., no published data show that the virus can transform and immortalize cells), the presence of HHV-8 infection has now been unquestionably linked to KS. There is, however, increasing evidence in patients with established KS that no single serologic assay detects evidence for HHV-8 infection in all patients *(2,3)*, whereas certain methods detect antibodies in a substantial minority of unaffected, low-risk individuals *(2,3)*. Clearly, therefore, the sensitivity and specificity of these tests need to be improved, and the relevant serologic responses must be identified. Studies are also needed to confirm whether antibody titer correlates with subsequent KS risk *(4)* and to evaluate whether detection or burden of HHV-8 DNA in cells adds predictive value to the serologic assays that are better suited to large-scale screening.

If we accept that not all individuals with evidence of HHV-8 infection are at equal risk for KS development, we also need to ask whether other factors interact with HHV-8 either to promote KS development or to protect against it. Those who treat KS have often observed sudden, rapid tumor progression around the time an opportunistic infection develops and improvement when the infection was treated, lending credence to the idea that excess production of inflammatory cytokines as well as increased HIV replication itself may stimulate KS growth. Also observed, on less frequent occasions, has been regression of KS lesions when active antiretroviral therapy was instituted *(5)*. These observations in patients with established KS are the clinical counterpart to numerous laboratory observations predating the dis-

**Affiliation of author:* Clinical Immunology Service, Division of Hematologic Oncology, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY.

Correspondence to: Susan E. Krown, M.D., Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, NY 10021.

© Oxford University Press

covery of HHV-8 that implicated a wide range of cytokines, angiogenic factors, and HIV products in KS pathogenesis *(6–9)*. We now need to look more closely at the role of these factors as codeterminants, with HHV-8, of future KS development. Within the AIDS Malignancy Consortium (AMC), a recently established, National Institutes of Health-supported multicenter trials group, active discussions are under way to design these predictive investigations and to use them to develop prophylaxis and natural history trials.

Pathogenesis-Based Therapeutic and Prophylactic Strategies for KS

Several cytotoxic agents have been identified that can induce regression, albeit temporary, of established, advanced KS *(10)*. At the National AIDS Malignancy Conference, the activity of two recent introductions to the therapeutic armamentarium, paclitaxel and vinorelbine, was described *(11,12)*. Paclitaxel's ability to inhibit angiogenesis *(13)* may partially explain its apparent high activity and the relatively long duration of response it induces in patients with advanced, refractory disease. Although the introduction of these agents increases the options for treatment of aggressive, widespread KS, there is still a need for therapy that induces higher and longer-lasting response rates. There is also a perceived need on the part of many investigators to develop effective therapy that makes better "biologic sense." This sort of targeted approach, based on the rapidly growing but still incomplete knowledge of KS pathogenesis, has tremendous intellectual appeal and has inspired a variety of clinical trials. With some exceptions, the results thus far have been mostly quite limited. However, for a number of these approaches directed against the formation of new blood vessels, inflammatory cytokines, and the viruses implicated either indirectly or directly in KS pathogenesis, there are some glimmers of activity that merit a closer look.

Antiviral Strategies

Both HIV-1 and HHV-8 are potential targets for anti-KS therapy. Although HIV-1 infection is not a requirement for development or progression of KS, active infection with HIV-1 is associated with high levels of inflammatory cytokines implicated in the development of KS lesions *(14,15)*. In addition, the HIV-1 Tat protein, which is released into the extracellular milieu by productively infected cells, also serves as a mitogen for KS-derived spindle cells *in vitro* and acts synergistically with growth factors and inflammatory cytokines to increase KS cell proliferation *(8,9)*. It should not be surprising, therefore, that effective control of HIV-1 replication achieved through the use of highly active antiretroviral drug regimens has now been reported to induce KS regression in some patients *(5)* and to obviate the need for long-term maintenance chemotherapy in others *(16)*. It is also possible that the improved KS response rates seen when nucleoside reverse transcriptase inhibitors are added to interferon alfa *(17,18)* may be partially attributable to the synergistic antiretroviral activity of these agents. Anecdotal reports that established KS regressed after treatment with foscarnet *(19)* and several studies that showed a decreased incidence of subsequent KS in patients treated with antiherpesvirus agents *(20,21)* suggest that, under some circumstances, inhibition of HHV-8 may also be of therapeutic or prophylactic value. However, none of the agents tested is specific for HHV-8 (the treatments were directed against cytomegalovirus and the studies were not designed to evaluate effects on HHV-8), so the mechanism of KS inhibition in these cases is entirely speculative. It is also possible that part of interferon's anti-KS activity may involve direct HHV-8 inhibition. Although this possibility has not been specifically tested, interferon has been shown to inhibit a closely related virus, *Herpesvirus saimiri (22)*. Even if KS is the product of viral transformation, a consistent therapeutic effect of antiherpesvirus agents against established KS is probably unlikely, since the growth of transformed cells is often independent of the virus. It may be more plausible to envision a prophylactic role for such agents if they are administered prior to a putative transforming event.

Cytokine Inhibition Strategies

A number of proinflammatory cytokines whose production is reportedly increased in some patients with HIV infection have also been considered targets for KS therapy. Tumor necrosis factor (TNF) and interleukin 1 (IL-1) have each been shown to stimulate KS cell proliferation *in vitro*, either directly or through induction of interleukin 6 (IL-6) *(6,7)*, and various agents that modify their production or action have either been tested in early clinical trials or proposed as candidate therapeutic agents. Of particular note are the retinoids, which are multifunctional agents among whose many attributes are down-regulation of both IL-6 receptors and IL-6 production in myeloma cells *(23)*, and thalidomide, a specific inhibitor of TNF *(24)*. One retinoid, 9-*cis*-retinoic acid, which interacts with both the retinoic acid receptor and the retinoid X receptor, has shown anti-KS activity when applied locally to KS lesions in gel form *(25)*. The AMC is now testing this as a systemic, orally administered drug, and there are some early indications of anti-KS activity. In addition, Bernstein et al. *(26)* have recently presented preliminary evidence of anti-KS activity for an intravenously administered liposomal preparation of all-*trans*-retinoic acid. Preliminary clinical evidence from two studies *(27,28)* suggests that orally administered thalidomide may also have substantial anti-KS activity. Interferon alfa, which has demonstrated activity against KS in some patients *(29)*, has been shown in clinical trials in normal volunteers and patients with chronic hepatitis to induce the production of the soluble TNF receptor and the IL-1 receptor antagonist *(30,31)*, which are endogenous down-regulators of these KS stimulatory cytokines. Other potential candidate agents that target cytokines, but which have not yet been tested, include the soluble TNF receptor, the IL-1 receptor antagonist, and the matrix metalloproteinase inhibitor marimistat, which also possesses inhibitory activity against TNF convertase, an enzyme that converts TNF to its active form.

Angiogenesis Inhibitory Strategies

A large number of strategies aimed at the inhibition of angiogenesis have been proposed. Some of these approaches have not yet been tested clinically, but candidate agents are already in development. These include, for example, inhibitors of matrix metalloproteinases, enzymes that facilitate capillary budding and invasion *(32)*, agents targeted at distinctive endothelial cell surface molecules present on proliferating tumor-associated vascu-

lature *(33,34)*, and agents that interfere with angiogenic polypeptide signal transduction *(35)*.

Agents utilizing other approaches have been tested with mixed results. For example, the fumagillin analogue TNP-470 has been studied in two phase I trials in patients with KS, yielding a few partial tumor responses in one study *(36)* and a significant reduction in tumor-associated edema in some patients *(37)*. Several heparin-binding compounds that are thought to inhibit angiogenesis by blocking basic fibroblast growth factor (bFGF) receptor binding have also been studied in phase I trials, and while the high hopes and expectations for these agents have not been realized, evidence for biologic activity has been observed. For example, in some patients treated with tecogalan, a significant reduction in KS-associated edema was seen *(38)*, perhaps signifying an important biologic effect. Interferon alfa and thalidomide are also capable of acting as angiogenesis inhibitors: Interferon inhibits bFGF gene expression and production in certain human tumor cell lines *(39)* and also inhibits endothelial cell motility, whereas thalidomide (in addition to its inhibitory effects on TNF production) may inhibit angiogenesis by multiple mechanisms, including inhibition of bFGF-induced vascular proliferation, intercellular adhesion, basement membrane formation, and overall vascular maturation *(40)*. The relative success of interferon alfa in controlling KS and possibly also thalidomide—although it is probably too early to comment on thalidomide's usefulness—may be a function of the sheer multiplicity of their potential targets and suggests that some of the other agents with more restricted targets may be successful in controlling KS only when used as part of combination regimens directed at multiple steps in the development of KS lesions. It is disappointing, therefore, to see that the further development of a number of these agents is not being pursued, at least partially, because of their apparent lack of substantial efficacy as single agents. However, as discussed by Judah Folkman at this conference, it may be that newer agents such as angiostatin *(41)* or endostatin *(42)* will be far more active than some of those already shown to have only marginal single-agent activity.

Effect of Improved HIV Therapy on KS Treatment Strategy

Patients with HIV infection are surviving longer as a result of improved antiretroviral therapy. These advances in HIV management are likely to benefit patients at all stages of HIV infection, including those with KS. The prospect of managing KS over a period of many years, rather than over a much shorter time period, raises several new questions and challenges that will need to be addressed when devising therapeutic plans and designing new therapeutic trials:

1) When in the course of KS should treatment begin and with which approaches or agents? Controversy about this issue preceded the introduction of HIV protease inhibitors, and the apparent benefits of state-of-the-art antiretroviral therapy on KS in some patients have made definition of the optimum KS treatment strategy even more complex.
2) A corollary of the first question is: Has more effective anti-HIV therapy rendered the need to find better KS therapy obsolete? Alternatively, are we merely in a "honeymoon" period that will end—as most honeymoons do—with increasing numbers of patients becoming affected by KS later in the course of their infection?
3) If we assume that the honeymoon will end, we need to ask: Will better control of HIV and its nonmalignant complications allow pathogenesis-based strategies to work better, and should this prompt a re-examination of some agents that were considered ineffective in the era before protease inhibitors?
4) We also need to ask: Will better control of HIV and its complications also lead to higher response rates, longer response durations, or a reduced need for maintenance therapy in patients on standard KS therapies?
5) Will the increasingly complex drug regimens of HIV-infected patients lead to untoward interactions with anti-KS therapies? Recently, for example, D. Scadden (personal communication, 1997) observed severe toxicity from paclitaxel when a patient who had been tolerating the drug well while on lamivudine, stavudine, and indinavir was switched to a combination of delavirdine, didanosine, and saquinavir. The metabolism of many cytotoxic agents is dependent on hepatic enzymes whose activity may be induced or inhibited by protease inhibitors and non-nucleoside reverse transcriptase inhibitors. Interferon alfa may also inhibit certain cytochrome P450 enzymes *(43)*, potentially affecting the metabolism of certain protease inhibitors. These potential drug interactions will be studied in several upcoming trials in patients with KS.
6) Do we need to be more concerned about the long-term consequences of chemotherapy for KS? There has been a recent trend toward the earlier use of chemotherapeutic agents like Doxil and DaunoXome. These drugs are well tolerated by most patients and rarely induce hair loss, nausea, or vomiting *(44,45)*, making many patients and perhaps even some physicians view them as something other than chemotherapy. Now that the life span of some patients with advanced KS may be measured in years, however, we need to be more concerned about the development of secondary cancers and other late complications of cytotoxic agents.
7) Have we gone as far as we can go with standard chemotherapeutic agents? Currently, few new cytotoxic agents are considered as candidates for testing in KS. The results of studies using combinations of several active drugs, such as the recently completed AIDS Clinical Trials Group (ACTG) comparison trial of Doxil alone or combined with bleomycin and vincristine *(46)* suggest that we may have reached a point of diminishing returns, where responses are not significantly increased but quality of life is diminished. Therefore, it is more important than ever to develop targeted, pathogenesis-based therapeutic strategies for KS.

Methods to Evaluate Benefits From KS Treatment

In order to measure the benefits of treatment for KS, it is necessary to document not only whether tumor regression has occurred, but also whether the patient is better and has shown tangible benefits that extend beyond tumor shrinkage. This is true for any tumor but may be especially so for KS, which may not always be directly life-threatening but may adversely affect quality of life by causing pain, disfigurement, or functional disability. The standardized methods of evaluation and criteria for response, developed and later refined by the Oncology Commit-

tee of the ACTG *(47)*, were an improvement over the vague and varied criteria used previously. However, these evaluation and response criteria have been criticized because of difficulties in the ability to reproduce and document lesion counts and measurements, the lack of standardized assessments of hard-to-measure visceral disease and edema, and the failure of the criteria to capture the clinical benefits such as pain relief, improvement in disfiguring lesions, and functional improvements induced by successful KS therapy.

Some of the problems with the ACTG criteria have stemmed from the failure of some investigators to consistently perform the repeated lesion counts and lesion measurements that were required to document response, but there is also general agreement that more rigorous standards of documentation are desirable. Feigal et al. *(48)* have described the joint efforts of the National Cancer Institute, the Food and Drug Administration, and the AMC to strengthen the process by which KS therapy is evaluated. This will include documentation not only of lesion size, character, and number in the skin and visceral sites, but also of physician and patient assessment of tumor-associated symptom relief in a function-oriented, quantitative way. Rigorous collection of the varied types of documentation required for these assessments will be tedious and time consuming but is an essential step in making effective therapy widely available to patients.

Summary and Conclusions

Several challenges must be met in order to improve the therapy of AIDS-associated KS over the coming years. The opportunity now exists to consider strategies to prevent KS development, but more sensitive and specific methods are needed to accurately identify those patients at highest risk. Increased knowledge of the biology of KS now permits the development of therapeutic strategies that target critical elements in KS pathogenesis. To some extent, these sorts of approaches have already been applied with some moderately successful examples; however, opportunities are expanding as a more sophisticated understanding of the complex processes involved in KS development unfolds and as an increasing number of agents are developed that influence specific pathogenetic targets. The introduction of highly active antiretroviral therapy now presents the opportunity to manage KS over a period of many years. Although more successful anti-HIV therapy may lead to fewer patients affected by KS or to a decrease in KS severity, it is also possible that longer survival with HIV infection may be associated with an increase in KS incidence over time. These possibilities raise important yet unanswered questions about optimum treatment strategies, the long-term consequences of cytotoxic chemotherapy, and the potential for adverse drug interactions as treatment regimens become more complex. Rigorous standards are needed to accurately measure the benefits of KS therapy, so that safe and effective new agents can be expeditiously introduced into clinical practice. Methods now exist to measure changes in tumor size and number, but development of quantitative scales to evaluate functional status, pain, and quality of life presents a more difficult challenge in the drug development process.

References

(1) Biggar RJ, Rabkin CS. The epidemiology of AIDS-related neoplasms. Hematol Oncol Clin North Am 1996;10:997–1010.

(2) Simpson GR, Schulz TF, Whitby D, Cook PM, Boshoff C, Rainbow L, et al. Prevalence of Kaposi's sarcoma associated herpesvirus infection measured by antibodies to recombinant capsid protein and latent immunofluorescence antigen. Lancet 1996;348:1133–8.

(3) Gao SJ, Kingsley L, Hoover DR, Spira TJ, Rinaldo CR, Saah A, et al. Seroconversion to antibodies against Kaposi's sarcoma-associated herpesvirus-related latent nuclear antigens before the development of Kaposi's sarcoma. N Engl J Med 1996;335:233–41.

(4) Ablashi DV, Chatlynne LG, Lapps W, Hoekzema DT, Masood R, Kramarsky B, et al. Human herpesvirus-8 (HHV-8): epidemiology and characterization of virus from KS-1 cell line. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A23.

(5) Routy JP, Urbanek A, MacLeod J, Thomas R. Significant regression of Kaposi's sarcoma following initiation of an effective antiretroviral combination treatment. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14: A22.

(6) Karp JE, Pluda JM, Yarchoan R. AIDS-related Kaposi's sarcoma. A template for the translation of molecular pathogenesis into targeted therapeutic approaches. Hematol Oncol Clin North Am 1996;10:1031–49.

(7) Miles SA. Pathogenesis of AIDS-related Kaposi's sarcoma. Evidence of a viral etiology. Hematol Oncol Clin North Am 1996;10:1011–21.

(8) Ensoli B, Barillari G, Salahuddin SZ, Gallo RC, Wong-Staal F. Tat protein of HIV-1 stimulates growth of cells derived from Kaposi's sarcoma lesions of AIDS patients. Nature 1990;345:84–6.

(9) Ensoli B, Gendelman R, Markham P, Fiorelli V, Colombini S, Raffeld M, et al. Synergy between basic fibroblast growth factor and HIV-1 Tat protein in induction of Kaposi's sarcoma. Nature 1994;371:674–80.

(10) Krown SE. Acquired immunodeficiency syndrome-associated Kaposi's sarcoma. Biology and management. Med Clin North Am 1997;81:471–94.

(11) Gill PS, Tulpule A, Reynolds T, Hadjenberg R, Mocharnuk R, Espina BM, et al. Paclitaxel (Taxol) in the treatment of relapsed or refractory advanced AIDS-related Kaposi's sarcoma [abstract]. Proc ASCO 1996;15:306.

(12) Errante D, Spina M, Nasti G, Tavio M, Vacchar E, Landonio G, et al. Evidence of activity of vinorelbine (VNR) in patients with previously treated epidemic Kaposi's sarcoma (KS). J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A36.

(13) Belotti D, Vergani V, Drudis T, Borsotti P, Pitelli MR, Viale G, et al. The microtubule-affecting drug paclitaxel has antiangiogenic activity. Clin Cancer Res 1996;2:1843–9.

(14) Ammann AJ, Palladino MA, Volberding P, Abrams D, Martin NL, Conant M. Tumor necrosis factors alpha and beta in acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. J Clin Immunol 1987;7: 481–5.

(15) Breen EC, Rezai AR, Nakajima K, Beall GN, Mitsuyasu RT, Hirano T, et al. Infection with HIV is associated with elevated IL-6 levels and production. J Immunol 1990;144:480–4.

(16) Volm MD, Wernz J. Patients with advanced AIDS-related Kaposi's sarcoma (EKS) no longer require systemic therapy after introduction of effective antiretroviral therapy [abstract]. Proc ASCO 1997;16:46a.

(17) Krown SE, Gold JW, Niedzwiecki D, Bundow D, Flomenberg N, Gansbacher B, et al. Interferon-α with zidovudine: safety, tolerance, and clinical and virologic effects in patients with Kaposi's sarcoma associated with the acquired immunodeficiency syndrome (AIDS). Ann Intern Med 1990;112: 812–21.

(18) Fischl MA, Finkelstein DM, He W, Powderly WG, Triozzi PL, Steigbigel RT, et al. A phase II study of recombinant human interferon-α2a and zidovudine in patients with AIDS-related Kaposi's sarcoma. AIDS Clinical Trials Group. J Acquir Immune Defic Syndr Hum Retrovirol 1996;11: 379–84.

(19) Morfeldt L, Torssander J. Long-term remission of Kaposi's sarcoma following foscarnet treatment in HIV-infected patients. Scand J Infect Dis 1994;26:749–52.

(20) Jones JL, Hanson DL, Chu SY, Ward JW, Jaffe HW. AIDS-associated Kaposi's sarcoma [letter]. Science 1995;267:1078–9.

(21) Mocroft A, Youle M, Gazzard B, Morcinek J, Halai R, Phillips AN. Antiherpesvirus treatment and risk of Kaposi's sarcoma in HIV infection. Royal Free/Chelsea and Westminster Hospitals Collaborative Group. AIDS 1996; 10:1101–5.

(22) Barahona HH, Melendez LV. *Herpesvirus saimiri: in vitro* sensitivity to virus-induced interferon and to polyriboinosinic acid polyribocytidylic acid. Proc Soc Exp Biol Med 1971;136:1163–7.

(23) Sidell N, Taga T, Hirano T, Kishimoto T, Saxon A. Retinoic acid-induced growth inhibition of a human myeloma cell line via down-regulation of IL-6 receptors. J Immunol 1991;146:3809–14.

(24) Moreira AL, Sampaio EP, Zmuidzinas A, Frindt P, Smith KA, Kaplan G. Thalidomide exerts its inhibitory action on tumor necrosis factor alpha by enhancing mRNA degradation. J Exp Med 1993;177:1675–80.
(25) Duvic M, Friedman-Kien AE, Miles SA, Looney DJ, Stevens V, Truglia J, et al. Phase I-II evaluation of Panretin™ (ALRT1057; LGD1057; AGN192013; 9-*cis*-retinoic acid) topical gel for AIDS-related cutaneous Kaposi's sarcoma [abstract]. Prog/Proc ASCO 1997;16:46a.
(26) Bernstein ZP, Cohen P, Rios A, Wagner B, Fischl M, Groopman J, et al. A multicenter, phase II/III study of Atragen™ (tretinoin liposomal) in patients with AIDS-associated Kaposi's sarcoma. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A19.
(27) Welles L, Little R, Wyvill K, Pluda J, Figg W, Lietzau J, et al. Preliminary results of a phase II study of oral thalidomide in patients with HIV infection and Kaposi's sarcoma (KS). J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A21.
(28) Bower M, Howard M, Gracie F, Phillips R, Fife K. A phase II study of thalidomide for Kaposi's sarcoma: activity and correlation with KSHV DNA load. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A35.
(29) Krown SE. Interferon and other biologic agents for the treatment of Kaposi's sarcoma. Hematol Oncol Clin North Am 1991;5:311–22.
(30) Tilg H, Mier JW, Vogel W, Aulitzky WE, Wiedermann CJ, Vannier E, et al. Induction of circulating IL-1 receptor antagonist by IFN treatment. J Immunol 1993;150:4687–92.
(31) Tilg H, Vogel W, Dinarello CA. Interferon-α induces circulating tumor necrosis factor receptor p55 in humans. Blood 1995;85:433–5.
(32) Ennis BW, Matrisian LM. Matrix degrading metalloproteinases. J Neurooncol 1993;18:105–9.
(33) Brooks PC, Montgomery AM, Rosenfeld M, Reisfeld RA, Hu T, Klier G, et al. Integrin $\alpha_v\beta_3$ antagonists promote tumor regression by inducing apoptosis of angiogenic blood vessels. Cell 1994;79:1157.
(34) Burrows FJ, Thorpe PE. Vascular targeting—a new approach to the therapy of solid tumors. Pharmacol Ther 1994;64:155–74.
(35) Kohn EC, Sandeen MA, Liotta LA. *In vivo* efficacy of a novel inhibitor of selected transduction pathways including calcium, arachidonate, and inositol phosphates. Cancer Res 1992;52:3208–12.
(36) Dezube BJ, vonRoenn JH, Holden-Wiltse J, Remick S, Cooley TP, Cheung TW, et al. Fumagillin analog (TNP-470) in the treatment of Kaposi's sarcoma: a phase I AIDS clinical trial group study. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A35.
(37) Pluda JM, Wyvill KK, Lietzau J, Figg WD, Whitcup SM, Feigel E, et al. A phase I trial administering the angiogenesis inhibitor TNP-470 (AGM-1470) to patients (pts) with HIV-associated Kaposi's sarcoma (KS). J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A19.
(38) Eckhardt SG, Burris HA, Eckardt JR, Weiss G, Rodriguez G, Rothenberg M, et al. A phase I clinical and pharmacokinetic study of the angiogenesis inhibitor, tecogalan sodium. Ann Oncol 1996;7:491–6.
(39) Singh RK, Gutman M, Bucana CD, Sanchez R, Llansa N, Fidler IJ. Interferons alpha and beta down-regulate the expression of basic fibroblast growth factor in human carcinomas. Proc Natl Acad Sci U S A 1995;92:4562–6.
(40) D'Amato RJ, Loughnan MS, Flynn E, Folkman J. Thalidomide is an inhibitor of angiogenesis. Proc Natl Acad Sci U S A 1994;91:4082–5.
(41) O'Reilly MS, Holmgren L, Shing Y, Chen C, Rosenthal RA, Moses M, et al. Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis lung carcinoma. Cell 1994;79:315–28.
(42) O'Reilly MS, Boehm T, Shing Y, Fukai N, Vasios G, Lane WS, et al. Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell 1997;88:277–85.
(43) Mannering GJ, Renton KW, El Azhary R, Deloria LB. Effect of interferon-inducing agents on hepatic cytochrome p-450 drug metabolizing systems. Ann NY Acad Sci 1980;350:314–31.
(44) Gill PS, Wernz J, Scadden DT, Cohen P, Mukwaya GM, vonRoenn JH, et al. Randomized phase III trial of liposomal daunorubicin versus doxorubicin, bleomycin, and vincristine in AIDS-related Kaposi's sarcoma. J Clin Oncol 1996;14:2353–64.
(45) Northfelt DW, Dezube BJ, Thommes JA, Levine R, vonRoenn JH, Dosik GM, et al. Efficacy of pegylated-liposomal doxorubicin in the treatment of AIDS-related Kaposi's sarcoma after failure of standard chemotherapy. J Clin Oncol 1997;15:653–9.
(46) Mitsuyasu R, vonRoenn J, Krown S, Kaplan L, Testa M. Comparison study of liposomal doxorubicin (Dox) alone or with bleomycin and vincristine (DBV) for treatment of advanced AIDS-associated Kaposi's sarcoma (AIDS-KS): AIDS Clinical Trial Group (ACTG) protocol 286 [abstract]. Prog/Proc ASCO 1997;16:5a.
(47) Krown SE, Metroka C, Wernz JC. Kaposi's sarcoma in the acquired immune deficiency syndrome: a proposal for uniform evaluation, response and staging criteria. J Clin Oncol 1989;7:1201–7.
(48) Feigal EG, vonRoenn J, Justice R, Yarchoan R, Krown S, Pluda J, et al. Kaposi's sarcoma response criteria: issues identified by the National Cancer Institute, Food and Drug Administration, and the AIDS Malignancy Consortium. J Acquir Immune Defic Syndr Hum Retrovirol 1997;14:A22.

Kaposi's Sarcoma-Associated Herpesvirus-Encoded Oncogenes and Oncogenesis

*Patrick S. Moore, Yuan Chang**

Molecular biologic studies of Kaposi's sarcoma-associated herpesvirus (KSHV) have identified a number of potential viral oncogenes that may contribute to KSHV-related neoplasia including a D-type cyclin, an IL-6-like cytokine, and a novel member of the interferon regulatory factor family. KSHV is functionally related to other DNA tumor viruses by encoding specific proteins to inhibit pRb, pro-apoptotic, and interferon-signaling tumor suppressor pathways. The virus appears to employ molecular piracy of cellular regulatory genes as a mechanism to avoid cellular antiviral responses. The transparency of the KSHV genome allows ready identification of the cellular regulatory pathways which may be involved in transformation by KSHV. This provides strong support to the notion that some tumor suppressor pathways serve the dual function of being antiviral pathways to induce cell cycle arrest, apoptosis, and enhanced cell-mediated immunity in response to virus infection. Neoplasia may result from specific viral strategies to overcome these host defense pathways. [Monogr Natl Cancer Inst 1998;23:65–71]

The second full year of working with Kaposi's sarcoma-associated herpesvirus (KSHV) is now finished and scientific studies on this virus are rapidly maturing. With the publication of the full-length genomic sequence *(1)*, new opportunities to investigate its molecular biology and virology are becoming available. This review will try to provide a road map for subsequent papers in this symposium on the major features of the KSHV genome and its potential oncogenes. Despite the early stage of KSHV molecular biologic studies, the virus is already providing unique insights into tumorigenesis and may serve as an important model for virus-induced oncogenesis.

The evidence gathered to date is overwhelming in support of KSHV as the infectious cause of Kaposi's sarcoma (KS) *(2)*. A review of 21 published studies involving 549 patients, with all forms of Kaposi's sarcoma from various parts of the world, shows that KSHV DNA can be detected on average 95% of the time in KS lesions by polymerase chain reaction. This exceeds the positivity rate for detection of papillomavirus in cervical cancer *(3)*. It is now clearly established using a variety of techniques that KSHV are not only localized to KS lesions *(4–6)*, but that viral DNA and RNA are detectable in most KS tumor spindle cells *(7–9)*.

Newly developed KSHV serologic assays also support a causal role for the virus in KS. As discussed by Weiss *(10)*, published assays show similar epidemiologic trends, despite important differences in their sensitivities and specificities. The relative KSHV seroprevalence measured by these assays matches the patterns for KS risk groups among patients with acquired immunodeficiency syndrome (AIDS) *(11–17)* and the patterns for non-HIV-related KS incidence among patients from various countries *(15,16,18)*. For KSHV to be causally related to KS, infection must precede onset of disease. This is an absolute criteria for causality, and both seroconversion *(13,18–20)* and DNA-based *(21,22)* detection studies provide evidence for KSHV infection among most AIDS-KS patients prior to disease onset.

Taken together, these and other studies *(23–26)* now indicate that KSHV is not a ubiquitous infection in most human populations, although high rates of infection in some populations [e.g., Italians and Ugandans *(18)*] may account for early studies suggesting widespread human KSHV infection *(27,28)*. It is thus likely that the virus is a necessary, but not necessarily sufficient cofactor for KS development. This does not diminish the role of KSHV in the genesis of KS and related neoplasias. With the possible exceptions of rabies and HIV, no infectious syndrome or illness is solely due to infection by a single agent without contributing host or environmental risk factors. Similarly, asymptomatic infection is the norm for most infectious agents and only a minority of persons infected with KSHV should be expected to eventually develop KS. This relatively simple view of KS causation has become complicated by the reeent finding of KSHV in cultured multiple myeloma stromal cells *(29)*. Whether or not this association with multiple myeloma, a cancer with a very different epidemiologic presentation from KS, is real remains an area for future research.

The important questions now become "How does KSHV cause KS, and what is the relative contribution of the virus to tumorigenesis vis-à-vis immunosuppression and other predisposing factors?" As indicated by Blackbourne et al. *(30)*, techniques have not yet been developed to allow *in vitro* high-titered virus transmission to other cell lines *(11,31)*. Therefore, much of the work on KSHV-related tumorigenesis and cell transformation has had to rely on molecular biologic investigation of individual genes. Like other herpesviruses, KSHV is presumed to replicate as a circular episome in latently infected cells and is

**Affiliations of authors:* P. S. Moore (School of Public Health), Y. Chang (Department of Pathology, College of Physicians and Surgeons), Columbia University, New York, NY.

Correspondence to: Patrick S. Moore, M.D., Department of Pathology, Columbia University, 630 W. 168th St., New York, NY 10032.

See "Notes" following "References."

© Oxford University Press

packaged in a linear genomic form during lytic replication. The importance of this is that standard dogma tells us lytic replication is uniformly fatal to the cell and that lytic phase gene expression cannot contribute to tumorigenesis. However, recent findings on lytic KSHV replication in KS lesions and the effects of viral DNA polymerase inhibitors on clinical disease suggest that we should carefully question this assumption *(8,32–34)*. As indicated by Ganem *(35)*, it is possible that a minority population of lytically infected cells is important in sustaining the KS lesion by constantly recruiting new, infected tumor cells directly or through paracrine mechanisms.

The KSHV genome was first determined by Renne et al. *(36)* to be approximately 165-kb long using purified virion-banded DNA *(36)*. This has been confirmed through genome mapping and sequencing studies *(1)*. We had previously estimated the genome size to be as large as 270 kb *(11)*, but this was determined using the BC-1 cell line which has a large genomic duplication *(1)*. The long unique region (LUR), approximately 140 kb in length, comprises the entire coding region for the virus and encodes at least 81 genes (Fig. 1).

Because of the high level of synteny between KSHV and herpesvirus saimiri (HVS, the prototypical member of the genus *Rhadinovirus*), KSHV genes are named after their corresponding HVS homologs starting from the left hand end of the genome. KSHV is missing the first three genes found in HVS *(37)* and the first homolog to a HVS gene begins at ORF 4. Synteny with HVS (except for ORFs 2 and 70) extends throughout the genome up to ORF 75 located on the right end of the LUR. Genes having detectable homology to HVS genes include structural protein and DNA synthetic enzyme genes which are conserved among all herpesviruses.

There are a number of genes, however, which are not found in HVS and are given a K designation (ORFs K1–15). No KSHV genes with sequence homology to the EBV Epstein-Barr nuclear antigens, latent membrane proteins, or the HVS saimiri transforming proteins involved in cell transformation are found in KSHV. Over 90% of the LUR (excepting approximately 12 kb on the right-hand portion of the genome) has also been recently sequenced from a KS lesion by Bernard Fleckenstein's group. While details of this sequencing project have not yet been published, comparison of the deposited KS lesion Genbank sequence to the BC-1 sequence *(1)* shows a remarkable degree of sequence conservation between BC-1 and KS isolates (generally <0.1% divergence), confirming previous comparisons *(11)*. Sequence divergence is most pronounced at the K1 locus and in internal repeat regions (e.g., Frnk, Vnet, Zppa, and Moi). The right end sequence has proven particularly difficult to clone into sequencing vectors *(1)* and it is unknown whether this part of the genome contains a hypervariable region similar to the left hand K1 locus.

Fig. 2 shows a simplified schematic diagram of the genome demonstrating the conserved herpesvirus gene blocks as first described by Chee et al. *(38)* from their sequencing of the cytomegalovirus genome. Genes that are conserved among members of all three herpesvirus subfamilies lie within these blocks. In between conserved gene blocks are regions that are unique to KSHV or are also found only in other rhadinoviruses. The

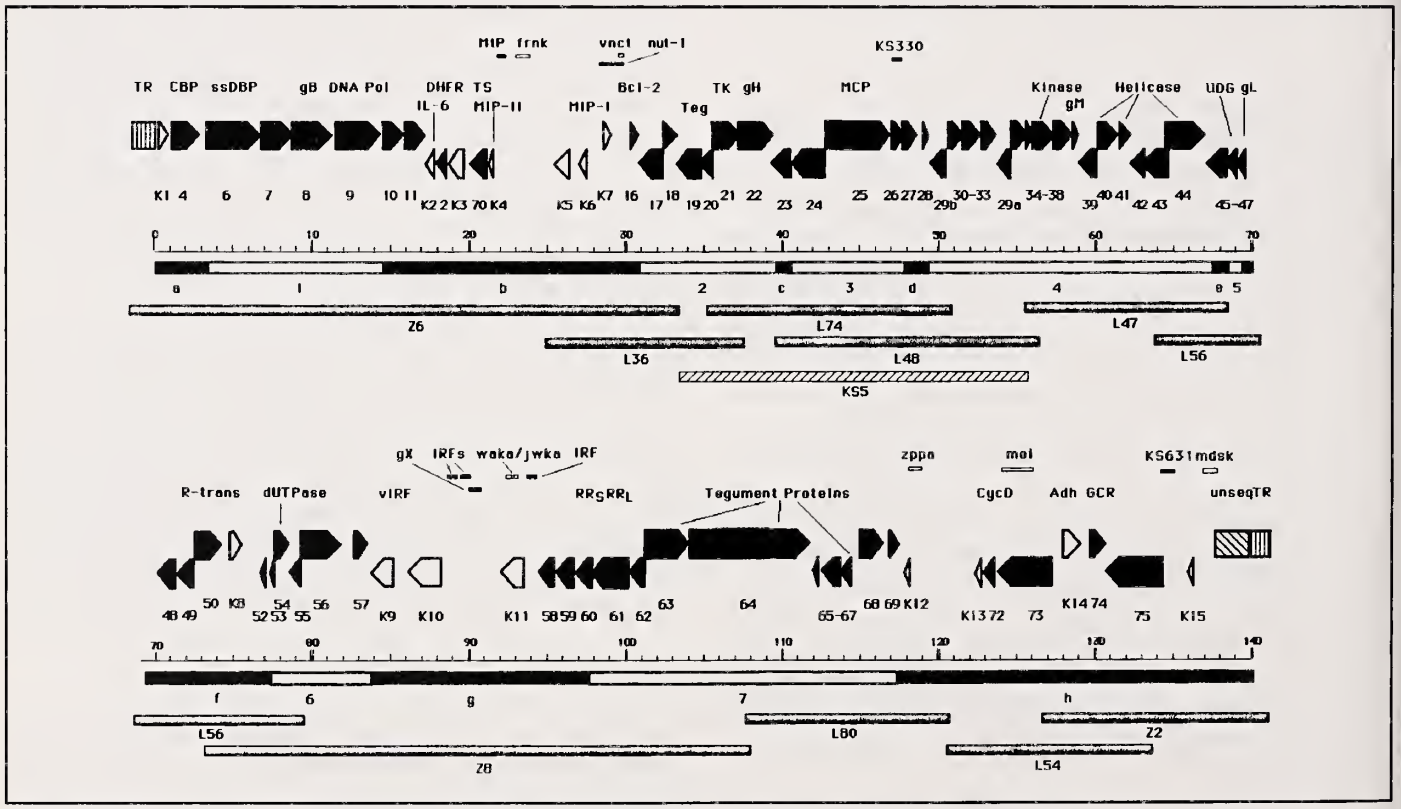


Fig. 1. The long unique region (LUR) of Kaposi's sarcoma herpesvirus (KSHV) is bracketted by high G:C content terminal repeat regions and contains the 81 genes thus far identified in the KSHV genome. KSHV genes are named after corresponding genes in herpesvirus saimiri (HVS) starting at the left end of the genome. Those genes without significant sequence homology are given a K prefix. The standard map of KSHV is in the same orientation as HVS *(37)* and is reversed compared to that of EBV *(95)*. Adapted from *(1)*.

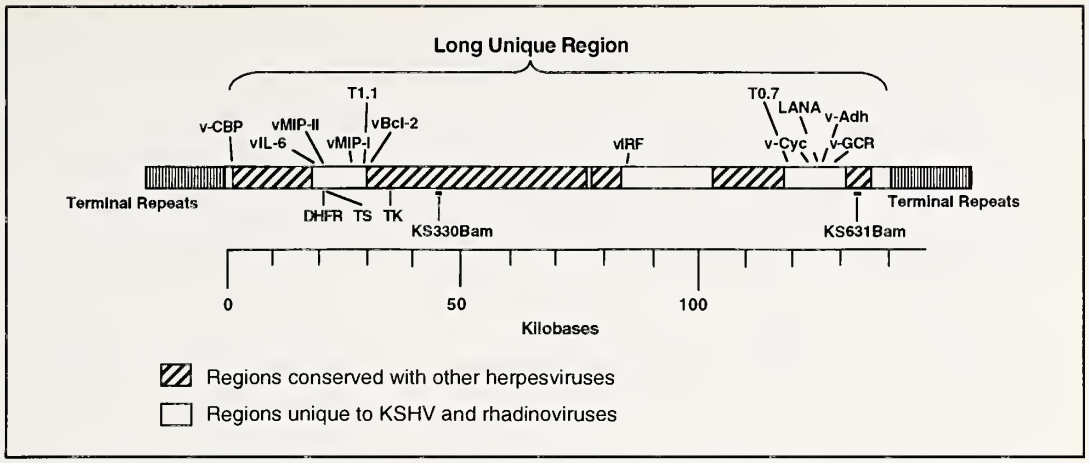


Fig. 2. A simplified representation of the Kaposi's sarcoma herpesvirus (KSHV) genome showing gene blocks conserved among herpesviruses and intervening regions without homology. KSHV homologs to cell regulatory and signaling protein genes are clustered within the nonconserved gene blocks.

unique regions contain a remarkable array of cellular gene homologs, encoding proteins involved in cell cycle regulation or cell signaling *(1,39)*. In addition, some DNA synthetic enzymes, such as dihydrofolate reductase (DHFR) and thymidylate synthase (TS), are encoded in these regions. Cellular homologs to the DNA synthetic enzymes encoded by the virus are under the control of the E2F transcriptional factor family *(40)*, suggesting that the virus has pirated those DNA synthesis genes that will allow it to replicate DNA outside of the S phase of the cell cycle. Two additional genes found in unique regions include the T1.1 and T0.7 (ORF K12) genes encoding polyadenylated transcripts *(41–43)* used for *in situ* hybridization studies because of their abundance in infected cells *(8)*.

What are the possible reasons for the virus having these cellular homologs? Our working hypothesis is that these viral proteins provide a defense against stereotypic cellular responses to viral infection. When a cell is infected by a virus it undergoes cell cycle shutdown, induction of apoptosis and *in vivo* enhancement of cell-mediated immunity through increased major histocompatibility antigen presentation [reviewed in *(44)*]. These effects may be in part mediated by retinoblastoma (pRb) and p53 tumor suppressor pathways, which are involved in the control of dysregulated cellular division *(45)* and, possibly, unregulated viral nucleic acid replication.

The cell cycle is regulated at the G_1 checkpoint by the E2F family of transcriptional factors that initiate transcription of a number of enzymes involved in DNA synthesis, allowing the cell to pass into the S phase of the cell cycle. Nonphosphorylated retinoblastoma protein binds to E2F preventing E2F directed transcription and stops the cell from progressing through the S phase of the cell cycle *(46,47)*. This can be abrogated by the activity of D-type cyclins interacting with various cyclin-dependent kinases, such as CDK6, which form a complex that phosphorylates pRb, causing release of active E2F, and allowing transcription of DNA synthesis genes.

p21 and related cyclin-dependent kinase inhibitors act as a counterbalance to the cyclin-CDK complex, preventing pRb phosphorylation and inhibition. Under conditions where retinoblastoma protein is mutated or when specific pRb inhibitors are present, such as adenovirus E1A, the cell can progress through the G_1 checkpoint in an uncontrolled fashion. A feedback mechanism exists *(48)*, however, such that overexpression of active E2F results in p53 activation that induces either transcription of p21 to reinitiate control of the cell cycle *(49)* or, failing this, induction of p53-mediated apoptosis *(50,51)* (Fig. 3).

This is a very simplified, and only to some degree correct, explanation of how these two tumor suppressor pathways interact with each other, but it is apparent that both tumor suppressor pathways are likely to act as antiviral pathways as well. A number of viruses have evolved specific gene products to inhibit both

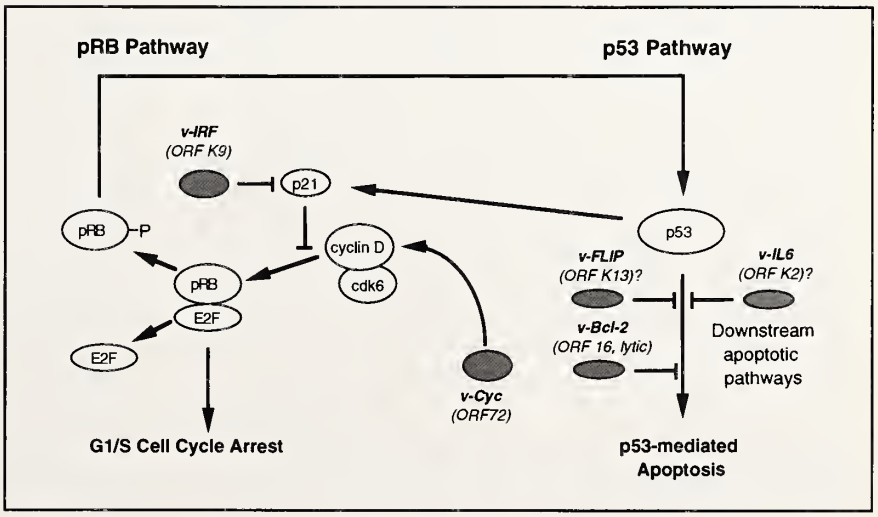


Fig. 3. Schematic diagram of the interaction between pRb and p53 tumor suppressor pathways that may serve the dual function of preventing successful cell colonization by a latently replicating DNA virus. Hypothetical sites of interaction with shaded Kaposi's sarcoma herpesvirus proteins are shown that may lead to dysregulated cell proliferation. Only limited experimental data are currently available to indicate that these viral proteins act as oncoproteins and several of these hypothetical interactions are based entirely on sequence homology to known genes alone.

the pRb and p53 activation *(52–54)*. Growth suppression imposed by pRb activity is a probable means of limiting replication of the naked DNA episomes of a latent, infecting virus. Adenovirus, for instance, encodes E1A which specifically inhibits pRb interactions with E2F to prevent G_0 arrest. However when E1A is expressed in an unopposed fashion, p53-mediated apoptosis is initiated. Adenovirus also encodes the antiapoptotic $E1B_{19k}$ and $E1B_{55k}$ proteins preventing p53-induced apoptosis. An important illustration of this counterbalance is seen with engineered adenovirus mutants that lack functional E1B genes and are only capable of replicative growth in p53-deficient tumor cells *(55)*. A similar, although phylogenetically and mechanistically distinct, interaction occurs for the papillomavirus E7 and E6 proteins with pRb and p53 pathways. Latent EBV gene expression may have a similar effect on these tumor suppressor pathways: work from Kieff's group *(56–58)* and others *(59)* demonstrates the potential antiapoptotic role of LMP1 in B cells, and data suggest that EBNA2 and EBNA-LP may interact to induce cyclin D2 overexpression *(60)*.

How does KSHV fit into this pattern? Like adenoviruses and papillomaviruses, KSHV also possesses a specific inhibitor of pRb function. KSHV directly encodes a homolog to D-type cyclins, v-cyc, on the right end of the genome at ORF72 *(61,62)*. ORF73, which encodes the major latency expressed antigen LANA *(9)*, and ORF74, which encodes an IL-8-like receptor that can induce cellular proliferation *(61,63)*, also lie in this region of the genome. v-cyc has been shown to be functionally active in phosphorylating pRb at authentic sites through interactions with cdk6 *(62)*, but it may also have broader specificity than cellular D-type cyclins in that it can mediate phosphorylation of histone H1 as well *(64,65)*. The altered substrate specificity of the viral cyclin suggests that it may act at other stages of the cell cycle in addition to the G_1/S checkpoint (Fig. 3).

With an active viral cyclin inhibiting pRb activity, one could also expect that there should be a counterbalancing antiapoptotic set of genes expressed by KSHV. Indeed, KSHV v-cyc overexpression in NIH3T3 cells rapidly induces cell death presumably through p53-mediated apoptosis (Boshoff C, Sarid R: unpublished observation). Not surprisingly, there are a number of KSHV genes that can potentially prevent cellular apoptosis. For example, the virus encodes a functional IL-6-like cytokine (ORF K2) which prevents B9 cell apoptosis *(39,66,67)* and is expressed only in KSHV hematopoietic cells, not KS tumor cells *(39)*. The vIL-6 appears to induce appropriate IL-6 Jak-STAT pathway activation [although its receptor specificity is different from hu-IL-6 *(68)*] and is thus likely to also induce antiapoptotic bcl-x_L protein production *(69)*.

KSHV also encodes a bcl-2 homologue (ORF16) simultaneously discovered by our group *(70)* and Cheng et al. *(71)* that has functional anti-apoptotic activity in both yeast and mammalian cells. There is disagreement as to whether or not v-bcl-2 heterodimerizes with cellular members of the bax-bcl-2 family, but it is clear that the KSHV v-bcl-2 prevents bax-mediated apoptosis. More recently, Thome et al. *(72)* have identified FLICE inhibiting proteins (v-FLIP) encoded by rhadinoviruses that possess dominant negative DEDD domains to inhibit apoptosis induced by CD95 pathway activation. No functional studies have been published yet on the corresponding KSHV protein (ORF K13), but it is an intriguing protein since it is expressed during viral latency in PEL and KS tumors, along with ORF72 and ORF73 (Sarid R, Wiezorek J, Moore PS, Chang Y: unpublished observation). It is not known whether or not v-FLIP has downstream activity that might cause inhibition of tumor suppressor pathway apoptosis. We are at an early stage in understanding the interplay of these KSHV proteins with known tumor suppressor pathways, but it is not unreasonable to suppose that these proteins either singly or in combination are active in preventing apoptosis induced by KSHV-inhibition of pRb.

In addition to cell cycle shutdown and apoptosis, enhanced immune recognition is another important cellular defense against virus infection. Many of these antiviral effects are coordinated at the cellular level through interferon signal regulation. We have found one KSHV protein involved in immune regulation, the v-IRF encoded by ORF K9 *(39)*, which may have a unique mechanism for allowing KSHV to escape immune surveillance *(73)*. This gene actually has a relatively low degree of sequence homology with the family of interferon regulatory factors (IRF) involved in positively or negatively regulating interferon signal transduction *(74)*.

Interferon-β binding at its receptor induces activation through the Jak-STAT signaling pathway with the formation of a trimeric complex of phosphorylated STAT1, STAT2, and p48 (or ISGF3γ) *(75,76)*. This complex, called interferon-stimulated gene factor 3 (ISGF3), translocates to the nucleus and binds to enhancer elements (ISREs) in promoters of interferon-stimulated genes. Interferon-induced gene transcription effects the phenotypic changes associated with interferon stimulation of cells, including increased MHC I transcription *(77,78)*, shut down of the cell cycle [through transcription of the cyclin-dependent kinase inhibitor p21 *(79–81)*], and also possibly p53-independent apoptosis *(82)*. Interferon also induces expression of the interferon regulatory factors IRF1 and IRF2 which also bind to ISRE. IRF1 probably initiates an amplifying loop that markedly amplifies the interferon signal whereas the more slowly degraded IRF2 shuts off this amplification cascade by competitive inhibition at ISRE sites. Thus, it is apparent that cellular proliferation control is modulated by interferon-mediated immune signaling *(74,83)*.

KSHV vIRF inhibits IFN-β signal transduction in a specific and dose-dependent fashion as measured by an ISRE-containing CAT reporter construct *(73)*. Further, vIRF transfection prevents IFNβ-induced transcription of the cyclin-dependent kinase inhibitor p21 suggesting a possible effect on cell cycle regulation. Expression of vIRF in NIH3T3 cells induces full cellular transformation and vIRF-expressing NIH3T3 cells form tumors when injected into nude mice. vIRF appears to have highly reproducible oncogenic activity in the NIH3T3 assay; however, these data should be interpreted to suggest only that this gene is active in cellular proliferation. It is unlikely that vIRF expression alone is responsible for KSHV-related human tumor formation and it is more likely to be only one component contributing to a multigenic process *(73)*.

Despite the early stage of molecular studies on KSHV, a functional comparison to other DNA tumor viruses is now reasonable (Table 1). Like other DNA tumor viruses, KSHV encodes specific proteins that are candidates for overcoming pRb-mediated growth arrest and p53-mediated apoptotic activity.

Table 1. Tumor suppressor pathway inhibition by selected tumor virus proteins

	pRB pathway	p53 pathway	IFN signaling pathway
Adenovirus	E1A	$E1B_{19k}$,$E1B_{55k}$	E1A (through p300)
Papillomavirus	E7	E6	E7?
Papovavirus (SV40)	LT	LT	LT (through p300)?
KSHV	v-Cyc (ORF72)	v-FLIP (ORF K13) v-IL-6 (ORF K2) v-bcl-2 (ORF16)	v-IRF (ORF K9)

Further, like the E1A protein of adenoviruses *(84,85)*, vIRF may specifically abrogate interferon-mediated responses, contributing to tumor induction as well. There is a close correlation between the cellular genes induced by EBV and the genes pirated by KSHV *(1)* suggesting that both herpesviruses modify their cellular environments in similar ways: KSHV brings in its own cellular homologs into the cell whereas EBV induces cellular genes to modify regulatory pathways essential to its survival. Whether additional common tumor virus pathways contributing to cell transformation will be found remains to be seen.

By examining the KSHV genome, a number of potential oncogenes have been identified by sequence homology alone. This does not necessarily mean that they function as transforming oncogenes *in vivo* and care is needed not to overinterpret the results of isolated gene studies. Primary effusion (or body cavity-based) lymphomas are clearly outgrowths of malignant clones *(86)* and provide a model for cell transformation caused in part, or entirely, by viral oncogene expression. KS, however, is a complex tumor whose origins remain disputed. In one view, KS may result from hyperplastic cellular proliferation driven by exogenous cytokines while an opposing view holds that it may result from KSHV oncogene-driven cellular proliferation. Recent HUMARA *(87,88)* and KSHV terminal repeat analyses *(1)* provide evidence for a clonal proliferation of tumor cells in KS lesions and support the latter supposition although these findings are not universally accepted *(89,90)*. This is further complicated by the fact that cell clonality does not necessarily result from virus-driven cell proliferation since transformed cells could originate from multiple virus-infected loci. In addition, KSHV encodes secreted cytokines *(39)* which are likely to play a role in non-neoplastic disorders, such as Castleman's disease *(91,92)*. This debate need not be polarizing, since it is likely that both paracrine and endogenous viral factors will ultimately be found to contribute to KS pathogenesis.

Additionally, *in vivo* expression of potential KSHV-encoded oncogenes is critical to understanding their contribution to these neoplastic disorders. Some genes, such as vIL-6, are expressed in infected hematopoietic tissues but not KS lesions. Other genes, such as the v-bcl-2 may only be expressed during lytic cycle replication *(70)* and are less likely to contribute to cellular transformation. Finally, isolated gene studies do not take into consideration immune surveillance, which we already know is likely to be tremendously important for this virus since KS is primarily a disease occurring in severely immunosuppressed patients.

Because of the early stage of KSHV, the list of oncogenes and their interactions is almost certainly incomplete. The growth-promoting ability of other viral proteins, such as the GPCR *(93)*, the MIP homologs *(94)*, and K1 (Jung J: personal communication), is an active area of investigation and promises a rich source of information on cell–virus interactions. Beyond providing pathogenic insights into KS and primary effusion lymphomas, KSHV is an important new model virus-induced tumorigenesis. KSHV appears to be another example of a tumor virus which is well-adapted to its host and causes tumors primarily in altered or "accidental" environments. Most tumor viruses cause tumors only under conditions where there is immunosuppression, complementing mutations in host cells, or during infection of non-native hosts. It is difficult to imagine the evolutionary benefit for the virus to cause a tumor in which it replicates latently and is therefore nontransmissible, which may kill the host. In the setting of severe AIDS-related immunosuppression, however, KSHV appears to be an extremely potent inducer of tumor formation. Early in the AIDS epidemic, up to 50% of gay male AIDS patients eventually developed KS *(93)* suggesting an extraordinarily high rate of tumor development among those persons who were both KSHV-infected and immunocompromised.

The convergent evolution of DNA tumor viruses to inhibit similar tumor suppressor pathways by different mechanisms suggests that these pathways also serve antiviral functions. Tumor viruses apparently require a mechanism for inhibiting tumor suppressor pathways if they are to survive as a latent episome in an actively dividing cell. It should not be surprising for cells to use tumor suppressor pathways to control both cancer cell growth and viral replication since both cases involve control of dysregulated DNA replication. Careful examination of KSHV and other tumor viruses will continue to lead to new insights into mechanisms of nonviral carcinogenesis.

Ready identification of regulatory gene homologs encoded by KSHV gives this virus a unique degree of transparency. Despite its complex interaction with cellular regulatory pathways, KSHV's option to use molecular piracy provides accessible starting points for examining these pathways. The extensive degree of symmetry between genes encoded by KSHV and those induced by EBV *(1)* also indicates that insights gained from one virus may to some degree be transferable to the other. KSHV illustrates in a new fashion that several major cellular pathways preventing tumor outgrowth, such as the p53 and pRb pathways, may also guard against latent virus infection. Virus strategies to overcome these intracellular defenses appear to inadvertently contribute to cell transformation. KSHV may thus provide a new, unique, and tractable model for exploring these and other issues in viral oncogenesis.

References

(1) Russo JJ, Bohenzky RA, Chien MC, Chen J, Yan M, Maddalena D, et al. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). Proc Natl Acad Sci U S A 1996;93:14862–7.

(2) Olsen SJ, Moore PS. Kaposi's sarcoma-associated herpesvirus (KSHV/HHV8) and the etiology of KS. In: Freidman H, Medveczky P, Bendinelli M, editors. Molecular immunology of herpesviruses. New York: Plenum Press. In press.
(3) IARC: Human Papillomaviruses. In: IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (IARC, ed). Lyon, France, 1995: 90–233.
(4) Lebbe C, Agbalika F, de CP, Deplanche M, Rybojad M, Masgrau E, et al. Detection of human herpesvirus 8 and human T-cell lymphotropic virus type 1 sequences in Kaposi sarcoma. Arch Dermatol 1997;133:25–30.
(5) Dupin N, Grandadam M, Calvez V, Gorin I, Aubin JT, Harvard S, et al. Herpesvirus-like DNA in patients with Mediterranean Kaposi's sarcoma. Lancet 1995;345:761–2.
(6) Moore PS, Chang Y. Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma lesions from persons with and without HIV infection. N Engl J Med 1995;332:1181–5.
(7) Boshoff C, Schulz TF, Kennedy MM, Graham AK, Fisher C, Thomas A, et al. Kaposi's sarcoma-associated herpesvirus infects endothelial and spindle cells. Nat Med 1995;1:1274–8.
(8) Staskus KA, Zhong W, Gebhard K, Herndier B, Wang H, Renne R, et al. Kaposi's sarcoma-associated herpesvirus gene expression in endothelial (spindle) tumor cells. J Virol 1997;71:715–9.
(9) Rainbow L, Platt GM, Simpson GR, Sarid R, Gao SJ, Stoiber H, et al. The 226- to 234-kilodalton latent nuclear protein (LNA) of Kaposi's sarcoma-associated herpesvirus (KSHV/HHV8) is encoded by ORF73 and is a component of the latency-associated nuclear antigen. J Virol 1997;71:5915–21.
(10) Boshoff C, Talbot S, Whitby D, Reeves J, Weiss RA. AIDS-associated tumors and HHV8. In: National AIDS Malignancy Conference. Bethesda (MD), 1997:A13.
(11) Neipel F, Albrecht JC, Fleckenstein B. Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? J Virol 1977;71:4187–92.
(12) Miller G, Rigsby MO, Heston L, Grogan E, Sun R, Metroka C, et al. Antibodies to butyrate-inducible antigens of Kaposi's sarcoma-associated herpesvirus in patients with HIV-1 infection. N Engl J Med 1996;334: 1292–7.
(13) Gao SJ, Kingsley L, Hoover DR, Spira TJ, Rinaldo CR, Saah A, et al. Seroconversion to antibodies against Kaposi's sarcoma-associated herpesvirus-related latent nuclear antigens before the development of Kaposi's sarcoma. N Engl J Med 1996;335:233–41.
(14) Kedes DH, Operskalski E, Busch M, Kohn R, Flood J, Ganem D. The seroepidemiology of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus): distribution of infection in KS risk groups and evidence for sexual transmission. Nat Med 1996;2:918–24.
(15) Simpson GR, Schulz TF, Whitby D, Cook PM, Boshoff C, Rainbow L, et al. Prevalence of Kaposi's sarcoma associated herpesvirus infection as measured by antibodies to recombinant capsid protein and latent immunofluorescence antigen. Lancet 1996;348:1133–8.
(16) Lennette ET, Blackbourne DJ, Levy JA. Antibodies to human herpesvirus type 8 in the general population and in Kaposi's sarcoma patients. Lancet 1996;348:858–61.
(17) Kedes DH, Ganem D, Ameli N, Bacchetti P, Greenblatt R. The prevalence of serum antibody to human herpesvirus 8 (Kaposi sarcoma-associated herpesvirus) among HIV-seropositive and high-risk HIV-seronegative women. JAMA 1997;277:478–81.
(18) Gao SJ, Kingsley L, Li M, Zheng W, Parravicini C, Ziegler J, et al. KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. Nat Med 1996;2:925–8.
(19) Andre S, Schatz O, Bogner JR, Zeichhardt H, Stoffler MM, Jahn HU, et al. Detection of antibodies against viral capsid proteins of human herpesvirus 8 in AIDS-associated Kaposi's sarcoma. J Mol Med 1997;75:145–52.
(20) Parravicini C, Olsen SJ, Capra M, Poli F, Sirchia G, Gao SJ, et al. Risk of Kaposi's sarcoma-associated herpesvirus transmission from donor allografts among Italian posttransplant Kaposi's sarcoma patients. Blood 1997; 90:2826–9.
(21) Whitby D, Howard MR, Tenant-Flowers M, Brink NS, Copas A, Boshoff C, et al. Detection of Kaposi's sarcoma-associated herpesvirus (KSHV) in peripheral blood of HIV-infected individuals predicts progression to Kaposi's sarcoma. Lancet 1997:364:799–802.
(22) Moore PS, Kingsley L, Holmberg SD, Spira T, Conley LJ, Hoover D, et al. Kaposi's sarcoma-associated herpesvirus infection prior to onset of Kaposi's sarcoma. AIDS 1996;10:175–80.
(23) Corbellino M, Bestetti G, Galli M, Parravicini C. Absence of HHV-8 in prostate and semen [letter]. N Engl J Med 1996;335:1238–9.
(24) Howard MR, Whitby D, Bahadur G, Suggett F, Boshoff C, Tenant FM, et al. Detection of human herpesvirus 8 DNA in semen from HIV-infected individuals but not healthy semen donors. AIDS 1997;11:F15–F19.
(25) Dupin N, Gorin I, Escande JP, Calvez V, Grandadam M, Huraux JM, et al. Lack of evidence of any association between human herpesvirus 8 and various skin tumors from both immunocompetent and immunosuppressed patients. Arch Dermatol 1997;133:537–7.
(26) Boshoff C, Talbot S, Kennedy M, O'Leary J, Schulz T, Chang Y. HHV8 and skin cancers in immunosuppressed patients [letter] [published erratum appears in Lancet 1996;348:138]. Lancet 1996;347:338–9.
(27) Monini P, de Lellis L, Fabris M, Rigolin F, Cassai E. Kaposi's sarcoma-associated herpesvirus DNA sequences in prostate tissue and human semen. N Engl J Med 1996;334:1168–72.
(28) Monini P, de Lellis L, Cassai E. Absence of HHV-8 in prostate and semen [letter]. N Engl J Med 1996;335:1238–9.
(29) Rettig MB, Ma HJ, Vescio RA, Pold M, Schiller G, Belson D, et al. Kaposi's sarcoma-associated herpesvirus infection of bone marrow dendritic cells from multiple myeloma patients. Science 1997;276:1851–4.
(30) Blackbourne DJ, Lennette E, Mourich D, Witte M, Way D, Nelson J, et al. Host range studies of human herpesvirus 8 (HHV8). In: National AIDS Malignancy Conference. Bethesda (MD), 1997:A34.
(31) Foreman KE, Friborg JJ, Kong WP, Woffendin C, Polverini PJ, Nickoloff BJ, et al. Propagation of a human herpesvirus from AIDS-associated Kaposi's sarcoma. N Engl J Med 1997;336:163–71.
(32) Jones J, Peterman T, Chu S, Jaffe H. AIDS-associated Kaposi's sarcoma. Science 1995;267:1078–9.
(33) Mocroft A, Youle M, Gazzard B, Morcinek J, Halai R, Phillips AN. Antiherpesvirus treatment and risk of Kaposi's sarcoma in HIV infection. AIDS 1996;10:1101–5.
(34) Herndier B, Shiramizu B, Weinstein M, Ganem D, Orenstein JM. KSHV latent and lytic infection in Kaposi's sarcoma lesions and lymphoid tissue. In: National AIDS Malignancy Conference. Bethesda (MD), 1997.
(35) Ganem D. Latent and lytic replication of HHV-8/KSHV. In: National AIDS Malignancy Conference. Bethesda (MD), 1997.
(36) Renne R, Lagunoff M, Zhong W, Ganem D. The size and conformation of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) DNA in infected cells and virions. J Virol 1996;70:8151–4.
(37) Albrecht JC, Nicholas J, Biller D, Cameron KR, Biesinger B, Newman C, et al. Primary structure of the *Herpesvirus saimiri* genome. J Virol 1992; 66:5047–58.
(38) Chee MS, Bankier SB, Bohni CM, Brown RC, Horsnell T, Hutchison CA, et al. Analysis of the protein coding content of the sequence of cytomegalovirus strain AD169. Curr Top Microbiol Immunol 1990;154:125–69.
(39) Moore PS, Boshoff C, Weiss RA, Chang Y. Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV. Science 1996; 274:1739–44.
(40) Adams PD, Kaelin WJ. Transcriptional control by E2F. Semin Cancer Biol 1995;6:99–108.
(41) Sun R, Lin SF, Gradoville L, Miller G. Polyadenylated nuclear RNA encoded by Kaposi sarcoma-associated herpesvirus. Proc Natl Acad Sci U S A 1996;93:11883–8.
(42) Zhong WD, Ganem D. Characterization of ribonucleoprotein complexes containing an abundant polyadenylated nuclear RNA encoded by Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8). J Virol 1997;71: 1207–12.
(43) Zhong W, Wang H, Herndier B, Ganem D. Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma. Proc Natl Acad Sci U S A 1996;93:6641–6.
(44) Shen Y, Shenk TF. Viruses and apoptosis. Curr Opin Genet Devel 1995; 5:105–11.
(45) Weinberg RA. The cat and mouse games that genes, viruses, and cells play. Cell 1997;88:573–5.
(46) Sherr CJ. D-type cyclins. Trends Biochemical Sci 1995;20:187–90.
(47) Sherr CJ. Cancer cell cycles. Science 1996;274:1672–7.
(48) White E. Tumour biology. p53, guardian of Rb. Nature 1994;371:21–2.
(49) Michieli P, Li W, Lorenzi MV, Miki T, Zakut R, Givol D, et al. Inhibition of oncogene-mediated transformation by ectopic expression of p21Waf1 in NIH3T3 cells. Oncogene 1996;12:775–84.
(50) Wu X, Levine AJ. p53 and E2F-1 cooperate to mediate apoptosis. Proc Natl Acad Sci U S A 1994;91:3602–6.
(51) Hiebert SW, Packham G, Strom DK, Haffner R, Oren M, Zambetti G, et al. E2F-1:DP-1 induces p53 and overrides survival factors to trigger apoptosis. Mol Cell Biol 1995;15:6864–74.
(52) White E. Regulation of p53-dependent apoptosis by E1A and E1B. Curr Top Microbiol Immunol 1995;199:34–58.
(53) Teodoro JG, Branton PE. Regulation of apoptosis by viral gene products. J Virol 1997;71:1739–46.
(54) Farthing AJ, Vousden KH. Functions of human papillomavirus E6 and E7 oncoproteins. Trends Microbiol 1994;2:170–4.
(55) Bischoff JR, Kirn DH, Williams A, Heise C, Horn S, Muna M, et al. An adenovirus mutant that replicates selectively in p53-deficient human tumor cells. Science 1996;274:373–6.

(56) Kieff E. Viral-mediated mechanisms of cancer pathogenesis. In: National AIDS Malignancy Conference. Bethesda (MD), 1997.
(57) Mosialos G, Birkenbach M, Yalamanchili R, VanArsdale T, Ware C, Kieff E. The Epstein-Barr virus transforming protein LMP1 engages signaling proteins for the tumor necrosis factor receptor family. Cell 1995;80: 389–99.
(58) Kieff E. Epstein-Barr virus—increasing evidence of a link to carcinoma [editorial]. N Engl J Med 1995;333:724–6.
(59) Okan I, Wang Y, Chen F, Hu LF, Imreh S, Klein G, et al. The EBV-encoded LMP1 protein inhibits p53-triggered apoptosis but not growth arrest. Oncogene 1995;11:1027–31.
(60) Sinclair AJ, Palmero I, Peters G, Farrell PJ. EBNA-2 and EBNA-LP cooperate to cause G0 to G1 transition during immortalization of resting human B lymphocytes by Epstein-Barr virus. EMBO J 1994;13:3321–8.
(61) Cesarman E, Nador RG, Bai F, Bohenzky RA, Russo JJ, Moore PS, et al. Kaposi's sarcoma-associated herpesvirus contains G protein-coupled receptor and cyclin D homologs which are expressed in Kaposi's sarcoma and malignant lymphoma. J Virol 1996;70:8218–23.
(62) Chang Y, Moore PS, Talbot SJ, Boshoff CH, Zarkowska T, Godden-Kent D, et al. Cyclin encoded by KS herpesvirus. Nature 1996;382:410.
(63) Arvanitakis L, Geras RE, Varma A, Gershengorn MC, Cesarman E. Human herpesvirus KSHV encodes a constitutively active G-protein-coupled receptor linked to cell proliferation. Nature 1997;385:347–50.
(64) Godden-Kent D, Talbot SJ, Boshoff C, Chang Y, Moore PS, Weiss RA, et al. The cyclin encoded by Kaposi's sarcoma associated herpesvirus (KSHV) stimulates cdk6 to phosphorylate the retinoblastoma protein and histone H1. J Virol 1997;71:4193–8.
(65) Li M, Lee H, Yoon DW, Albrecht JC, Fleckenstein B, Neipel F, et al. Kaposi's sarcoma-associated herpesvirus encodes a functional cyclin. J Virol 1997;71:1984–91.
(66) Nicholas J, Ruvolo VR, Burns WH, Sandford G, Wan X, Ciufo D, et al. Kaposi's sarcoma-associated human herpesvirus-8 encodes homologues of macrophage inflammatory protein-1 and interleukin-6. Nat Med 1997;3: 287–92.
(67) Neipel F, Albrecht JC, Ensser A, Huang YQ, Li JJ, Friedman KA, et al. Human herpesvirus 8 encodes a homolog of interleukin-6. J Virol 1997; 71:839–42.
(68) Molden J, Chang Y, Yow Y, Moore PS, Goldsmith MA. A Kaposi's sarcoma-associated herpesvirus-encoded cytokine homolog (vIL-6) activates signaling through the shared gp130 receptor subunit. J Biol Chem 1997;272:19625–31.
(69) Schwarze MM, Hawley RG. Prevention of myeloma cell apoptosis by ectopic bcl-2 expression or interleukin 6-mediated up-regulation of bcl-xL. Cancer Res 1995;55:2262–5.
(70) Sarid R, Sato T, Bohenzky RA, Russo JJ, Chang Y. Kaposi's sarcoma-associated herpesvirus encodes a functional Bcl-2 homologue. Nat Med 1997;3:293–8.
(71) Cheng EH, Nicholas J, Bellows DS, Hayward GS, Guo HG, Reitz MS, et al. A Bcl-2 homolog encoded by Kaposi sarcoma-associated virus, human herpesvirus 8, inhibits apoptosis but does not heterodimerize with Bax or Bak. Proc Natl Acad Sci U S A 1997;94:690–4.
(72) Thome M, Schneider P, Hofman K, Fickenscher H, Meinl E, Neipel F, et al. Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. Nature 1997;386:517–21.
(73) Gao SJ, Boshoff C, Jayachandra S, Weiss RA, Chang Y, Moore PS. KSHV ORF K9 (vIRF) is an oncogene that inhibits the interferon signaling pathway. Oncogene 1997;15:1979–86.
(74) Taniguchi T, Harada H, Lamphier M. Regulation of the interferon system and cell growth by the IRF transcription factors. J Cancer Res Clin Oncol 1995;121:516–20.
(75) Schindler C, Fu XY, Improta T, Aebersold R, Darnell JJ. Proteins of transcription factor ISGF-3: one gene encodes the 91-and 84-kDa ISGF-3 proteins that are activated by interferon alpha. Proc Natl Acad Sci U S A 1992;89:7836–9.
(76) Schindler C. Cytokine signal transduction. Receptor 1995;5:51–62.
(77) Weisz A, Kirchhoff S, Levi BZ. IFN consensus sequence binding protein (ICSBP) is a conditional repressor of IFN inducible promoters. Inter Immunol 1994;6:1125–31.
(78) Maudsley DJ, Pound JD. Modulation of MHC antigen expression by viruses and oncogenes. Immunol Today 1991;12:429–31.
(79) Hobeika AC, Subramaniam PS, Johnson HM. IFNα induces the expression of the cyclin-dependent kinase inhibitor p21 in human prostate cancer cells. Oncogene 1997;14:1165–70.
(80) Chin YE, Kitagawa M, Su WC, You ZH, Iwamoto Y, Fu XY. Cell growth arrest and induction of cyclin-dependent kinase inhibitor p21 WAF1/CIP1 mediated by STAT1. Science 1996;272:719–22.
(81) Sangfelt O, Erickson S, Einhorn S, Crander D. Induction of Cip/Kip and Ink4 cyclin dependent kinase inhibitors by interferon-alpha in hematopoietic cell lines. Oncogene 1997;14:415–23.
(82) Tanaka N, Ishihara M, Lamphier MS, Nozawa H, Matsuyama T, Mak TW, et al. Cooperation of the tumour suppressors IRF-1 and p53 in response to DNA damage. Nature 1996;382:816–8.
(83) Harada H, Kitagawa M, Tanaka N, Yamamoto H, Harada K, Ishihara M, et al. Anti-oncogenic and oncogenic potentials of interferon regulatory factors-1 and -2. Science 1993;259:971–4.
(84) Reich N, Pine R, Levy D, Darnell JJ. Transcription of interferon-stimulated genes is induced by adenovirus particles but is suppressed by E1A gene products. J Virol 1988;62:114–9.
(85) Bhattacharya S, Eckner R, Grossman S, Oldread E, Arany Z, D'Andrea A, et al. Cooperation of Stat2 and p300/CBP in signalling induced by interferon-alpha. Nature 1996;383:344–7.
(86) Cesarman E, Chang Y, Moore PS, Said JW, Knowles DM. Kaposi's sarcoma-associated herpesvirus-like DNA sequences are present in AIDS-related body cavity based lymphomas. N Engl J Med 1995;332:1186–91.
(87) Rabkin CS, Bedi G, Musaba E, Sunkutu R, Mwansa N, Sidransky D, et al. AIDS-related Kaposi's sarcoma is a clonal neoplasm. Clin Cancer Res 1995;1:257–60.
(88) Rabkin CS, Janz S, Lash A, Coleman AE, Musaba F, Liotta L, et al. Monoclonal origin of multicentric Kaposi's sarcoma lesions. N Engl J Med 1997;336:988–93.
(89) Diaz-Cano SJ, Wolfe HJ. Clonality in Kaposi's sarcoma. N Engl J Med 1997;337:571.
(90) Gill P, Tsai Y, Rao AP, Jones P. Clonality in Kaposi's sarcoma. N Engl J Med 1997;337:570–1.
(91) Soulier J, Grollet L, Oskenhendler E, Cacoub P, Cazals-Hatem D, Babinet P, et al. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. Blood 1995;86:1276–80.
(92) Parravicini C, Corbellino M, Paulli M, Magrini U, Lazzarino M, Moore PS, et al. Expression of a virus-derived cytokine, KSHV vIL-6, in HIV seronegative Castleman's disease. Am J Pathol. In press.
(93) Holmberg S, Buchbinder S, Conley L, Wong L, Katz M, Penley K, et al. The spectrum of medical conditions and symptoms before acquired immunodeficiency syndrome in homosexual and bisexual men infected with the human immunodeficiency virus. Am J Epidemiol 1995;141:395–404.
(94) Boshoff C, Endo Y, Collins PD, Takeuchi Y, Reeves JD, Schweickart VL, et al. Angiogenic and HIV-inhibitory functions of KSHV-encoded chemokines. Science 1997;278:290–4.
(95) Baer R, Bankier AT, Biggin PL, Deininger PL, Farrell PJ, Gibson TJ, et al. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature 1984;310:207–11.

Notes

Supported by Public Health Service grant CA67391 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

We thank the members of our laboratory who have contributed to the projects briefly described here, including Roy Bohenzky, Shou-Jiang Gao, Sonja Olsen, Ronit Sarid, and Jeffrey Wiezorek. We would also like to acknowledge the contributions of our collaborators, especially Drs. Robin Weiss and Chris Boshoff of the Institute for Cancer Research, James Russo and Izzy Edelman of the Columbia Genome Center, and Thomas Schulz from the University of Liverpool. We also thank Laurie Anderson for her help in preparing this manuscript.

Human Herpesvirus 8—the First Human *Rhadinovirus*

*Frank Neipel, Jens-Christian Albrecht, Bernhard Fleckenstein**

Kaposi's sarcoma (KS)-associated herpesvirus, also known as human herpesvirus 8 (HHV-8), is the first known human member of the genus *Rhadinovirus*. It is regularly found by polymerase chain reaction in all forms of KS, in certain types of Castleman's disease, and in body cavity-based B-cell lymphoma. Other members of this virus group occur in nonhuman primates, ungulates, rabbits, and mice and cause in part fulminant lymphomas and other neoplastic disorders of the hematopoietic system. Rhadinoviruses share a typical genome structure; most characteristically, they contain numerous sequences that appear to be sequestered from cellular DNA. We cloned and sequenced almost the complete genome of HHV-8 from a single KS biopsy specimen. Although this procedure revealed collinear organization and extensive homologies with the open reading frames of herpesvirus saimiri, genes with homology to the known oncoproteins (Stp, Tip) were not identified in the HHV-8 genome. However, HHV-8 reading frame K1, the positional analogue of Stp/Tip, was found to be significantly variable between different strains. We found, in addition, the reading frames for homologues of cellular interleukin 6, macrophage inflammatory proteins α and β (MIP1α and MIP1β, respectively), an interferon-responsive factor, and two inhibitors of apoptosis. Several of these cell-homologous genes of HHV-8 have already been shown to code for functional proteins. [Monogr Natl Cancer Inst 1998;23:73–77]

Human herpesvirus 8 (HHV-8), also referred to as Kaposi's sarcoma (KS)-associated herpesvirus (KSHV), is the first known human member of the genus *Rhadinovirus (1)*. Members of this group of herpesviruses share a common genome structure: A central segment of low-GC DNA (L-DNA) is flanked by multirepetitive high-GC DNA (H-DNA) *(2–6)* (Fig. 1).

In addition to their common genome structure, which is the cause of physical properties reflected by the term "rhadinovirus" (ῥαδινός [Greek] = fragile), all animal rhadinoviruses known so far share a common epidemiology. They are frequent in their natural host (>50%), where infection is not known to be associated with apparent disease (Table 1). In contrast, infection of closely related species is often the cause of fulminant lymphoproliferative diseases or overt malignant lymphoma with fatal outcome. For example, there is no indication for pathogenic properties of herpesvirus saimiri in its natural host, the squirrel monkey. However, in related species (e.g., common marmosets and several other New World primates), infection by herpesvirus saimiri results in polyclonal T-cell lymphomas *(7)*. Similarly, alcelaphine herpesvirus type 1 and the related ovine herpesvirus type 2 are nonpathogenic in wildebeest and sheep, respectively, whereas a lymphoproliferative syndrome termed "malignant catarrhal fever" is caused upon infection of cattle *(4)*. This finding does not necessarily imply that HHV-8 would be exceptional if it is causing a malignant tumor in its natural host. The incidence of KS, body cavity-based lymphomas, and multicentric forms of Castleman's disease is so low that such disease associations would have certainly been overlooked in animal models. Likewise, Epstein-Barr virus (EBV), which is the closest known relative of HHV-8 in humans, is clearly associated with infectious mononucleosis and with several cancers. However, primary EBV infection is usually not apparent unless additional factors, such as immunosuppression or delay of primary infection, disturb the delicate balance of virus and host. Infection of species other than the well-adapted natural host is possibly another example of an "accident" associated with increased tumorigenicity.

Thus, despite the lack of clear-cut evidence for oncogenesis by nonhuman rhadinoviruses in their natural hosts, the finding that infection of non-natural hosts is frequently associated with lymphoproliferative syndromes hints at the pathogenic potential that these viruses might show even in their natural hosts under certain, albeit unusual, circumstances. It has been shown for herpesvirus saimiri that the gene(s) relevant for malignant transformation are located close to the left end of the genome (Fig. 2) *(8–11)*. A single reading frame termed "STP-A" is present there in herpesvirus saimiri subgroup A viruses. This reading frame is required for the transforming phenotype of herpesvirus saimiri subgroup A, as shown by deletion mutants *(12)*. Two reading frames, termed "STP-C" and "TIP," are present at the equivalent position of herpesvirus saimiri subgroup C viruses, and both are obviously related to the oncogenic phenotype *(10,13)*. Expression of STP-A in transgenic mice results in lymphoid tumors *(14)*, whereas animals transgenic for STP-C develop epithelial tumors *(15)*. In addition, subgroup C strains of herpesvirus saimiri can transform human T lymphocytes *(16)*. These cells still depend on interleukin 2, but the respective T-cell antigen is no longer required for continuous growth in cell culture [reviewed in *(7)*]. Sequencing the complete genome of HHV-8 from both a body cavity-based lymphoma cell line *(5)* and a single KS biopsy specimen has revealed that no obvious homologue of Stp/Tip or other viral oncogenes is encoded by HHV-8

**Affiliation of authors:* Institut für Klinische und Molekulare Virologie, Universität Erlangen-Nürnberg, Schloßgarten 4, Erlangen, Germany.

Correspondence to: Bernhard Fleckenstein, M.D., Institut für Klinische und Molekulare Virologie, Universität Erlangen-Nürnberg, Schloßgarten 4, D-91054 Erlangen, Germany. E-mail: fleckenstein@viro.med.uni-erlangen.de

See "Note" following "References."

© Oxford University Press

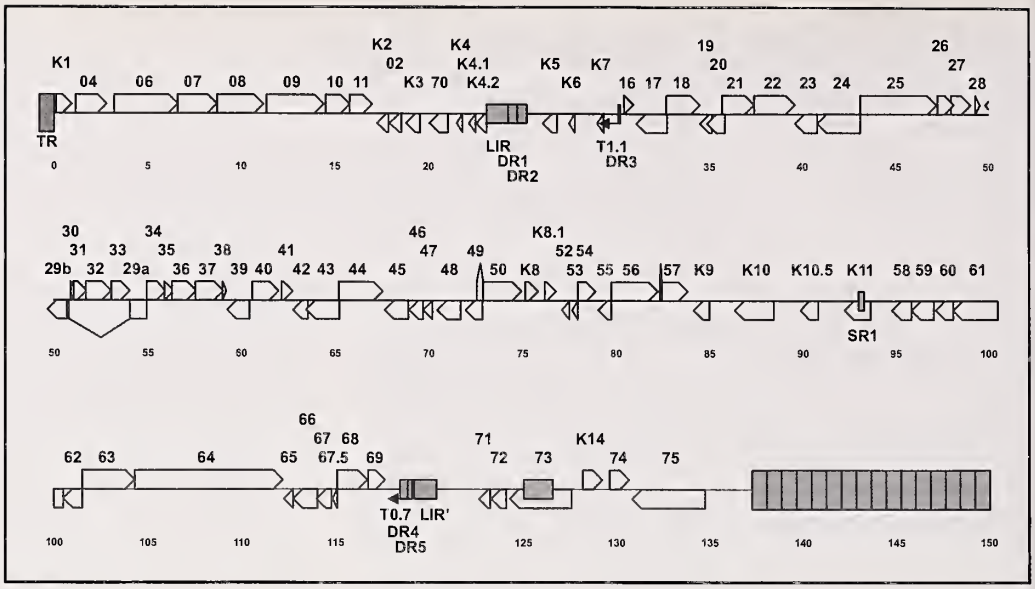


Fig. 1. Human herpesvirus 8 (HHV-8) genome organization. Arrowheads indicate orientation of identified open reading frames in the L-DNA segment of the HHV-8 genome. Open reading frames with homologues in the herpesvirus saimiri genome are numbered from 04 to 75; genes unique in HHV-8 are termed "K1 to K14." Regions of repetitive sequence are denoted by shaded boxes. TR = terminal repeat; LIR and LIR′ = long inverted repeat; DR1–DR5 = internal direct repeat; T1.1 = 1.1-kb transcript; T0.7 = 0.7-kb transcript.

(5,17). Instead, a reading frame, termed "K1," is present at the same position in HHV-8 (Figs. 1 and 2). This reading frame has no detectable sequence homology to STP or other transformation relevant genes. However, K1 has both a transmembrane region and a short intracytoplasmic tail like the transforming genes of other rhadinoviruses that are encoded at equivalent genomic positions. K1 is thus certainly a candidate for a transforming gene. Comparison of the two available complete HHV-8 genomic sequences reveals a striking degree of conservation. With the possible exception of the very right end of the unique region, reading frame K1 is by far the most divergent reading frame. The high degree of interstrain variability is another feature that K1 shares with the STP genes of herpesvirus saimiri. Although there is less than 0.1% divergence between the two complete sequences, 6% of the amino and nucleic acids are different between the K1 sequences derived from a KS and a B-cell line. This makes K1 a prime candidate for studies of HHV-8 molecular epidemiology.

Acquisition of genes from the host cell genome is a common feature of most herpesviruses and of rhadinoviruses in particular. So far, there are least 14 reading frames of HHV-8 that are clearly homologous to known cellular genes (Table 2). In contrast to their cellular counterparts, these viral genes are usually transcribed into an unspliced messenger RNA. This implies that not genomic DNA fragments but complementary DNA (cDNA) molecules were integrated into the viral genome. Reverse transcriptase activity must therefore have been present. One may speculate that not only was reverse transcription enabled by co-infection of the same cell by a retrovirus, but also cellular genes were transferred to the herpesvirus via retroviral genomes with their capability of integration. Although different members of the genus *Rhadinovirus* acquired different genes from the host cell, these genes are usually found at strikingly similar genomic positions (Fig. 2). These sequestered genes are frequently clustered around areas of repetitive sequence: close to the genomic termini and—most notably—in two genomic areas about 20–30 kilobase pairs apart from the ends of the unique region (Fig. 2). The pattern of repetitive elements in these areas includes stretches rich in AT (adenosine–thymidine) flanked by elements of dyad symmetry. These are the typical hallmarks of origins of DNA replication, although experimental data to confirm this are not available for any of the rhadinoviruses. Therefore, one scenario for the acquisition of foreign genes by a herpesvirus is linked to DNA replication. It could well be that, during replication by the rolling circle mechanism, fragments arise and, upon re-circularization, cDNA molecules are inserted.

Table 1. Pathogenic properties of rhadinoviruses in natural and foreign hosts

Virus type	Natural host	Pathogenic properties in natural host	Foreign hosts	Pathogenic properties in foreign hosts
Human herpesvirus type 8	Human	Putative: Kaposi's sarcoma, B-cell lymphomas		
Herpesvirus saimiri	Squirrel monkey		Marmosets, etc.	Polyclonal malignant T-cell lymphomas
Herpesvirus ateles	Spider monkey		Marmosets, etc.	Polyclonal malignant T-cell lymphomas
Alcelaphine herpesvirus type 1	Wildebeest		Cattle, buffalo	Malignant catarrhal fever
Ovine herpesvirus type 2	Sheep		Cattle	Malignant catarrhal fever
Bovine herpesvirus type 4	Cattle			
Herpesvirus sylvilagus	Cottontail rabbit	Benign lymphoproliferation		
Equine herpesvirus type 2	Horse	Foals: transient immunosuppression		
Murine herpesvirus-68	Bank vole		Mouse	Benign splenomegaly

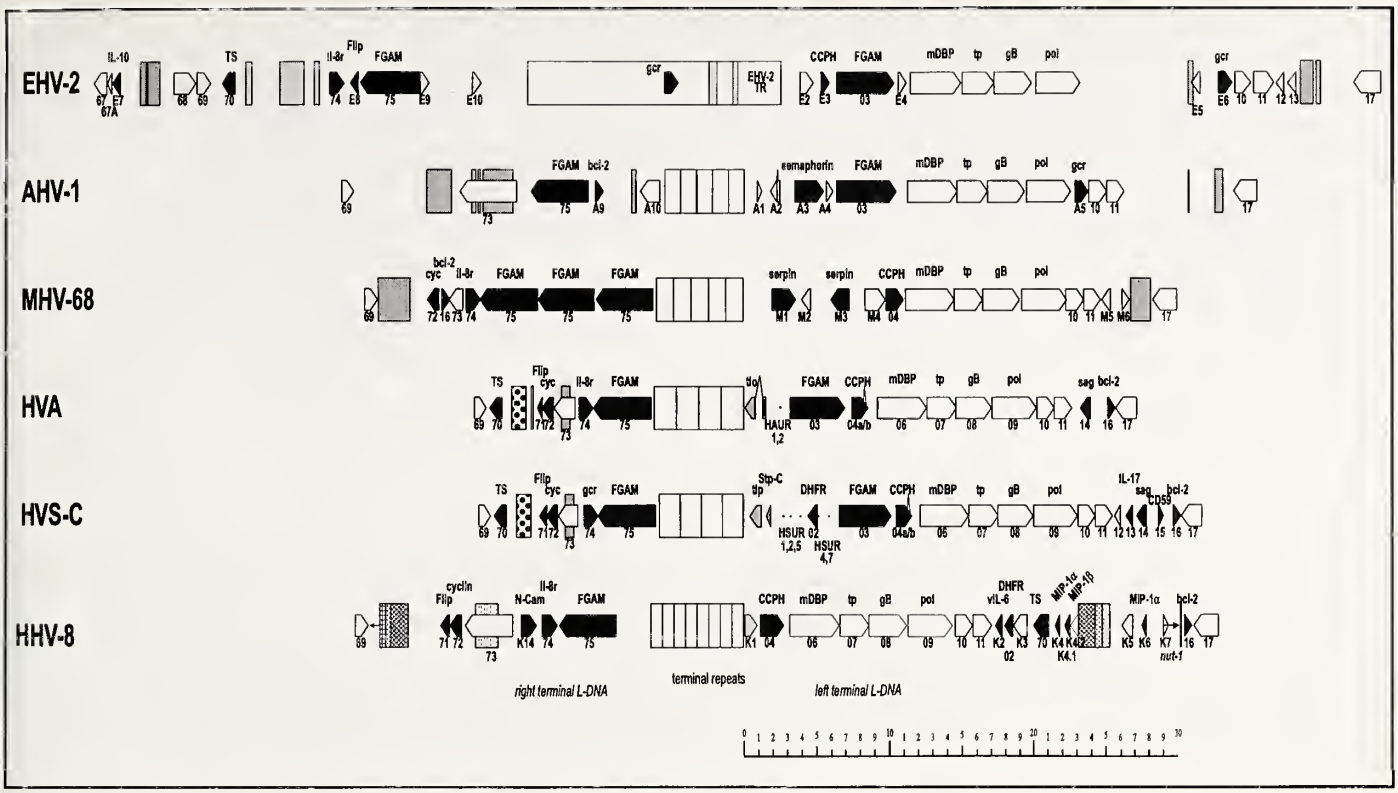


Fig. 2. Positions of cell-homologous genes in circular *Rhadinovirus* genomes. The termini of the L-DNA segments of genomes of HHV-8 (human herpesvirus-8), HVS-C (herpesvirus saimiri subgroup C), HVA (herpesvirus ateles), AHV-1 (alcelaphine herpesvirus type 1) *(4)*, EHV-2 (equine herpesvirus type 2), and MHV-68 (murine herpesvirus type 68) *(3)* are shown as oriented in circularized latent genomes. Cell-homologous genes are symbolized by dark gray arrows. Open arrows indicate reading conserved herpesvirus reading frames (mDBP = major DNA binding protein; tp = transport protein; gB = glycoprotein B; pol = DNA polymerase). Most reading frames with homology to known cellular genes are located close to the terminal repeats and in a nonconserved region flanked by ORF11 on one side and ORF17 on the other side. Repetitive elements (indicated by hatched boxes) are also present in this area. In addition to the genes shown here, an interferon response factor homologue K9 is encoded by HHV-8 in a nonconserved area between open reading frames 57 and 58 (Fig. 1). CCPH = complement control protein homologue; sag = open reading frame 14 of herpesvirus saimiri and herpesvirus ateles with homology to mouse mammary tumor virus superantigen; CD59 = viral CD59 homologue; DHFR = dihydrofolate reductase; TS = thymidylate synthase; cyc = viral cyclin D homologue; bcl-2 = viral bcl-2 homologue; Flip = Flice inhibitory protein; vIL-6 = viral interleukin 6; MIP1α = macrophage inflammatory protein α; MIP1β = open reading frame with homology to macrophage inflammatory protein 1β and macrophage chemoattractant protein; IL-8R = viral interleukin 8 receptor; IL-17 = viral interleukin 17; gcr = G-protein coupled receptor α/β; FGAM = phosphoribosylformylglycinamide synthase; HSUR = herpesvirus saimiri U-like RNA; HAUR = herpesvirus ateles U-like RNA.

Although for most of these captured genes it is not very likely that they are essential for virus replication in cell culture, they certainly have important functions in the viruses' natural habitat. Although different rhadinoviruses acquired different host-cell genes, their putative functions apparently converge to achieve three common goals: 1) to enhance DNA replication independently from the cell cycle, 2) to expand the pool of infectable cells, and 3) to counteract the host's responses to infection *(18)* (Table 3).

The first function, enhancement of DNA replication independent from the status of the infected cell, is achieved by enzymes of the nucleotide metabolism, i.e., thymidylate synthase, dihydrofolate reductase, and formylglycinoamidine synthase. Virus-encoded cyclins, interleukin 6, and the interleukin 8 receptor may enhance cell proliferation and may thus expand the pool of infectable cells. The three macrophage inflammatory proteins encoded by HHV-8 might work in a similar way by attracting susceptible cells. Apoptosis is a typical response of the host to infection by a virus. HHV-8 carries two genes, ORF16 (vbcl-2) and ORF71 (vFLIP), both of which could extend the life span of infected cells through the inhibition of apoptosis by two different mechanisms *(19–21)*. Similarly, the complement control protein homologues present in most rhadinoviruses counteract the host's response *(2,22)*. The virus-encoded interferon-response factor homologue (vIRF) might fit into this scenario at two different places: 1) It could counteract interferon-mediated suppression, and 2) it could mimic the proliferative effect of human interferon response factor 2. One can easily imagine how these genes can contribute to malignant growth transformation. Increasing the pool of available nucleotides not only enhances viral DNA replication but also facilitates the proliferation of transformed cells (Table 3). Genes that extend the pool of infectable cells or that prolong their life span, i.e., the virus-encoded cyclins, apoptosis inhibitors, cytokines, cytokine receptors, and interferon response factors, could also contribute to dysregulated growth and favor the development of cancers. Although under normal circumstances the functions of these viral genes are well balanced with the host, malignant growth does not occur. Like several other rhadinoviruses, HHV-8 has a host of genes that could accidentally contribute to tumor development. Examples of such accidents that favor malignant growth are infection of a non-natural host (Table 1), infection that occurs usually late in the host's life, and certainly infection of an immunosuppressed host. The initially mentioned finding that rhadinoviruses are not usually pathogenic in their natural host is therefore still compatible with the likelihood of HHV-8 being associated with at least two can-

Table 2. Cellular homologues of rhadinoviruses and the γ_1 herpesvirus Epstein-Barr virus (EBV)*

	HHV-8	HVS	HVA	AHV-1	BHV-4†	EHV-2	MHV-68	EBV
CCPH	+	+	+	−	−	−	+	−
vIL-6	+	−	−	−	−	−	−	−
DHFR	+	+	−	−	−	−	−	−
TS	+	+	+	−	−	+	−	−
CC-chemokines	3	−	−	−	−	−	−	−
bcl-2 family protein	+	+	+	+	+	−	+	+
Interferon	+	−	−	−	−	−	−	−
Responsive factor vFLIP	+	+	+	−	+	+	−	−
D-type cyclin	+	+	+	−	−	−	+	−
N-CAM family protein	+	−	−	−	−	−	−	−
IL-8R/gcr	+	+	+	+	−	3	+	−
FGAM	+	2	2	2	+	2	3	+
U-RNAs	−	≤7	2	−	−	−	−	−
Tyrosin kinase interacting protein	−	+/−	+	−	−	−	−	−
Protein with collagen motifs	−	+	+	−	−	−	−	−
IL-17	−	+	−	−	−	−	−	−
CD59	−	+	−	−	−	−	−	−
IL-10	−	−	−	−	n.k.	+	−	+
Semaphorin	−	−	−	+	−	−	−	−
Serpin	−	−	−	−	−	−	2	−

*CCPH = complement control protein homologue; vIL-6 = viral interleukin 6; DHFR = dihydrofolate reductase; TS = thymidylate synthase; vFLIP = viral Flice inhibitory protein; N-CAM family = neuronal cell-adhesive membrane protein family; IL-8R/gcr = interleukin 8 receptor or other G-protein-coupled receptor homologue; FGAM = phosphoribosylformylglyeinamidine synthase; IL-17 = interleukin 17 homologue; CD59 = CD59 homologue; IL-10 = interleukin 10 homologue; HHV-8 = human herpesvirus-8; HVS = herpesvirus saimiri; HVA = herpesvirus ateles; AHV-1 = alcelaphine herpesvirus type 1; BHV-4 = bovine herpesvirus type 4; EHV-2 = equine herpesvirus type 2; MHV-68 = murine herpesvirus 68. +/− = homologous reading frame present/absent or number of reading frames if more than one is present. n.k. = not known, not present at equivalent position. Numbers in columns represent multiple copies within the genomes.

†Positional conserved regions sequenced.

cers in humans. Diseases as infrequent as KS, body cavity-based lymphomas, and Castleman's disease would likely have been overlooked in animal models. All rhadinoviruses are highly pathogenic only if certain rare "accidents" coincide with infection, and infection of a non-natural host by animal rhadinoviruses can be seen as one example of such an accident, revealing the facultative pathogenic potential inherent to the rhadinoviruses.

Table 3. Putative functions of cell-homologous human herpesvirus 8 (HHV-8) genes*

ORF	Gene product	In natural lytic or persistent infection	In Kaposi's sarcoma, B-cell lymphoma and CD
		A) Cytokines and cytokine receptors	
K2	vIL-6	Growth stimulation of infected cells	Paracrine/autocrine growth stimulation of spindle/B cells
K4	vMIP1α	Chemotaxis of persistently infected hematopoietic cells	Chemoattraction of persistently infected cells
K4.1	vMIP1β	"	"
K5	vMIP1α	"	"
74	vGCR	Amplification of natural habitat	Dysregulated growth stimulation by constitutive action
		B) Genes involved in regulation of cell growth or survival	
K1	K1	?	Positional homologue of rhadinoviral transforming oncoproteins
04	CCPH	Protection against elimination by complement	Protection against elimination by complement
16	vbcl-2	Extension of life span of infected cells	Extension of life span of growth-transformed cells
K9	vIRF	Counteracting IFN-mediated virus suppression	Mimicking transforming effect of IRF-2 or interfering with antiproliferative action of IFN
71	vFLIP	Extension of life span of infected cells	Extension of life span of transformed cells
72	k-cyclin	Amplification of natural habitat	Dysregulated growth stimulation
		C) Genes involved in nucleotide metabolism	
02	vDHFR	Increasing nucleotide pool for virus replication	Increasing nucleotide pool for cell proliferation
70	vTS	"	"
75	vFGAM-synthase	"	"

*ORF = open reading frame; vIL-6 = viral interleukin 6; vMIP1α/β = viral macrophage inflammatory protein α/β; vGCR = viral G-protein-coupled receptor; CCPH = complement control protein homologue; vIRF = viral interferon response factor homologue; vFLIP = viral Flice inhibitory protein; k-cyclin = cyclin D homologue of HHV-8/Kaposi's sarcoma-associated herpesvirus; vDHFR = viral dihydrofolate reductase; vTS = viral thymidylate synthase; vFGAM-synthase = viral phosphoribosylformylglycinamidine synthase; IRF-2 = interferon response factor-2; IFN = interferon.

References

(1) Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 1994;266:1865–9.

(2) Albrecht JC, Nicholas J, Cameron KR, Newman C, Fleckenstein B, Honess RW. Herpesvirus saimiri has a gene specifying a homologue of the cellular membrane glycoprotein CD59. Virology 1992;190:527–30.

(3) Telford EA, Watson MS, Aird HC, Perry J, Davison AJ. The DNA sequence of equine herpesvirus 2. J Mol Biol 1995;249:520–8.

(4) Ensser A, Pflanz R, Fleckenstein B. Primary structure of the alcelaphine herpesvirus 1 genome. J Virol 1997;71:6517–25.

(5) Russo JJ, Bohenzky RA, Chen MC, Chen J, Yan M, Maddalena D, et al. Nucleotide sequence of the Kaposi's sarcoma-associated herpesvirus (HHV8). Proc Natl Acad Sci U S A 1996;93:14862–7.

(6) Neipel F, Albrecht JC, Fleckenstein B. Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? J Virol 1997;71:4187–92.

(7) Meinl E, Fickenscher H, Fleckenstein B. Chemokine receptors and chemokine-inducing molecules of lymphotropic herpesviruses. Immunol Today 1996;17:199.

(8) Biesinger B, Trimble JJ, Desrosiers RC, Fleckenstein B. The divergence between two oncogenic herpesvirus saimiri strains in a genomic region related to the transforming phenotype. Virology 1990;176:505–14.

(9) Albrecht JC, Nicholas J, Biller D, Cameron KR, Biesinger B, Newman C, et al. Primary structure of the herpesvirus saimiri genome. J Virol 1992; 66:5047–58.

(10) Jung JU, Trimble JJ, King NW, Biesinger B, Fleckenstein BW, Desrosiers RC. Identification of transforming genes of subgroup A and C strains of herpesvirus saimiri. Proc Natl Acad Sci U S A 1991;88:7051–5.

(11) Biesinger B, Tsygankov AY, Fickenscher H, Emmrich F, Fleckenstein B, Bolen JB, et al. The product of the herpesvirus saimiri open reading frame 1 (tip) interacts with T cell-specific kinase p56lck in transformed cells. J Biol Chem 1995;270:4729–34.

(12) Desrosiers RC, Silva DP, Waldron LM, Letvin NL. Nononcogenic deletion mutants of herpesvirus saimiri are defective for *in vitro* immortalization. J Virol 1986;57:701–5.

(13) Duboise SM, Guo J, Czajak S, Desrosiers RC, Jung JU. STP and Tip are essential for herpesvirus saimiri oncogenicity. J Virol 1998;72:1308–13.

(14) Kretschmer C, Murphy C, Biesinger B, Beckers J, Fickenscher H, Kirchner T, et al. A herpes saimiri oncogene causing peripheral T-cell lymphoma in transgenic mice. Oncogene 1996;12:1609–16.

(15) Murphy C, Kretschmer C, Biesinger B, Beckers J, Jung J, Desrosiers RC, et al. Epithelial tumours induced by a herpesvirus oncogene in transgenic mice. Oncogene 1994;9:221–6.

(16) Biesinger B, Müller-Fleckenstein I, Simmer B, Lang G, Wittmann S, Platzer E, et al. Stable growth transformation of human T lymphocytes by herpesvirus saimiri. Proc Natl Acad Sci U S A 1992;89:3116–9.

(17) Neipel F, Albrecht JC, Ensser A, Huang YQ, Li JJ, Friedman Kien AE, et al. Primary structure of the Kaposi's sarcoma associated human herpesvirus 8. Genbank accession No. U93872.

(18) Moore PS, Boshoff C, Weiss RA, Chang Y. Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV. Science 1996; 274:1739–44.

(19) Cheng EH, Nicholas J, Bellows DS, Hayward GS, Guo HG, Reitz MS, et al. A Bcl-2 homolog encoded by Kaposi sarcoma-associated virus, human herpesvirus 8, inhibits apoptosis but does not heterodimerize with Bax or Bak. Proc Natl Acad Sci U S A 1997;94:690–4.

(20) Thome M, Schneider P, Hofmann K, Fickenscher H, Meinl E, Neipel F, et al. Viral FLICE-inhibitory proteins (FLIPs) prevents apoptosis induced by death receptors. Nature 1997;386:517–21.

(21) Sarid R, Sato T, Bohenzky RA, Russo JJ, Chang Y. Kaposi's sarcoma-associated herpesvirus encodes a functional bcl-2 homologue. Nat Med 1997;3:293–8.

(22) Fodor WL, Rollins SA, Bianco-Caron S, Rother RP, Guilmette ER, Burton WV, et al. The complement control protein homolog of herpesvirus saimiri regulates serum complement by inhibiting C3 convertase activity. J Virol 1995;69:3889–92.

Note

Supported by the Ria Freifrau von Firtsch Stiftung, the "Deutsche Krebshilfe—Dr. Mildred Scheel Stiftung" grant No. W134/94/FL2, and European Union grant BMH4-CT95-1016.

Novel Organizational Features, Captured Cellular Genes, and Strain Variability Within the Genome of KSHV/HHV8

*John Nicholas, Jian-Chao Zong, Donald J. Alcendor, Dolores M. Ciufo, Lynn J. Poole, Robert T. Sarisky, Chuang-Jiun Chiou, Xiaoqun Zhang, Xiaoyu Wan, Hong-Guang Guo, Marvin S. Reitz, Gary S. Hayward**

ing serologic and molecular probe correlations indicate that the newly discovered gamma herpesvirus KSHV or HHV8 is the likely etiologic agent of all forms of Kaposi's sarcoma as well as BCBL/PEL and MCD in patients with acquired immunodeficiency syndrome (AIDS). Two large segments of HHV8 DNA from an AIDS-associated BCBL tumor covering genomic positions 0–52 kilobase [kb] and 108–140 kb have been cloned, mapped, and partially sequenced. Our studies have focused on novel viral proteins encoded within a 13-kb divergent locus (DL-B) by nine capped homologues of cellular genes, including vIL-6, vDHFR, vTS, vBcl-2, three C-C beta chemokines (vMIP-1A, vMIP-1B, and vBCK), and two LAP/PHD subclass zinc finger proteins (IE1A and IE1B). The HHV-8 vIL-6, vDHFR, vTS, and vBcl-2 proteins have all been shown to be active in a variety of appropriate functional assays, and transcripts from vIL-6, vMIP-1B, vIE1-A, vIE1-B, and vDHFR genes are all expressed as abundant single messenger RNA species after butyrate or phorbol ester (TPA) induction of the lytic cycle in HHV8-positive BCBL cell lines. All of these genes lie within a divergent transcriptional domain that contains a single central enhancer and associated untranslated leader region plus seven distinct proximal promoters, some of which are negatively regulated through AP-1 and ZRE motifs by the EBV ZTA transactivator. This region also encompasses a predicted complex oriLyt domain of 1050 bp that is duplicated in inverted orientation adjacent to the T0.7 latency RNA in another large divergent locus (DL-E). We have previously described three distinct subtypes of the HHV8 genome that differ by 1.0%–1.5% at the nucleotide level within the ORF26 and ORF75 genes. Certain strains or clades appear to have preferential geographic distributions, but it is not known as yet whether there are any specific disease associations. Interestingly, the A, B, and C subtypes of HHV-8 proved to differ dramatically in coding content at both extreme left and right ends of the unique segment of the genome as well as in the positions of the junctions with the terminal repeats. On the left-hand side, the receptor-like ORF-K1 protein is highly variable with A-strain subtypes displaying 15% amino acid differences from C strains and up to 30% differences from B strains. On the right-hand side, two unrelated alternative types of the putative multiple membrane spanning ORF-K15 protein are found. [Monogr Natl Cancer Inst 1998:23:79–88]

Introduction

Genome Characteristics of Gamma Herpesviruses

The complete genomic DNA sequences of five distinct gamma herpesviruses are now available. A comparison of the organization and gene content of Epstein-Barr virus (EBV) *(1)*, herpesvirus saimiri (HVS) *(2)*, EHV-2 *(3)*, HHV-8 *(4)*, and MHV68 *(5)* is shown in Fig. 1. Major blocks (I, II, III, and IV) of conserved genes present in all herpesviruses are shown and individual conserved genes that have residual homology among all of the five gamma herpesvirus genomes are given as solid bars. Hatched or open bars show those genes that are present in only a subset of these viruses or are unique to individual viruses. These genes tend to group within six regions referred to as divergent loci (DL) A, B, C, D, E, and F. Prior to the discovery of HHV-8 *(6)*, the human B-cell trophic virus EBV and the squirrel monkey T-cell lymphoma-associated virus HVS were designated as the prototypes of two subgroups of gamma herpesviruses, namely, γ1 and γ2. A number of old-world primate viruses similar to EBV are also clearly γ1 viruses and several other new-world primate γ2 viruses (also known as Rhadinoviruses) are closely related to HVS. More recently, several equine (EHV-2), bovine (BHV-4, AHV-1), and murine (MHV68) viruses have also been tentatively assigned to the γ2 subclass based on data about their gene content and genome organization, although a case could be made for assigning EHV-2 to a separate subclass because of its higher GC content and the presence of

**Affiliations of authors:* J. Nicholas, J.-C. Zong, D. J. Alcendor, D. M. Ciufo, L. J. Poole, R. T. Sarisky, C.-J.- Chiou, X. Zhang, X. Wan, G. S. Hayward, Molecular Virology Laboratories, Oncology Center, The Johns Hopkins School of Medicine, Baltimore, MD; H.-G. Guo, M. S. Reitz, Institute of Human Virology, University of Maryland, Baltimore.

Correspondence to: Gary S. Hayward, Ph.D., Department of Pharmacology and Molecular Sciences, The Johns Hopkins University, 725 N. Wolfe St., WBSB 317, Baltimore, MD 21205. E-mail: Gary.Hayward@qmail.bs.jhu.edu

See "Notes" following "References."

© Oxford University Press

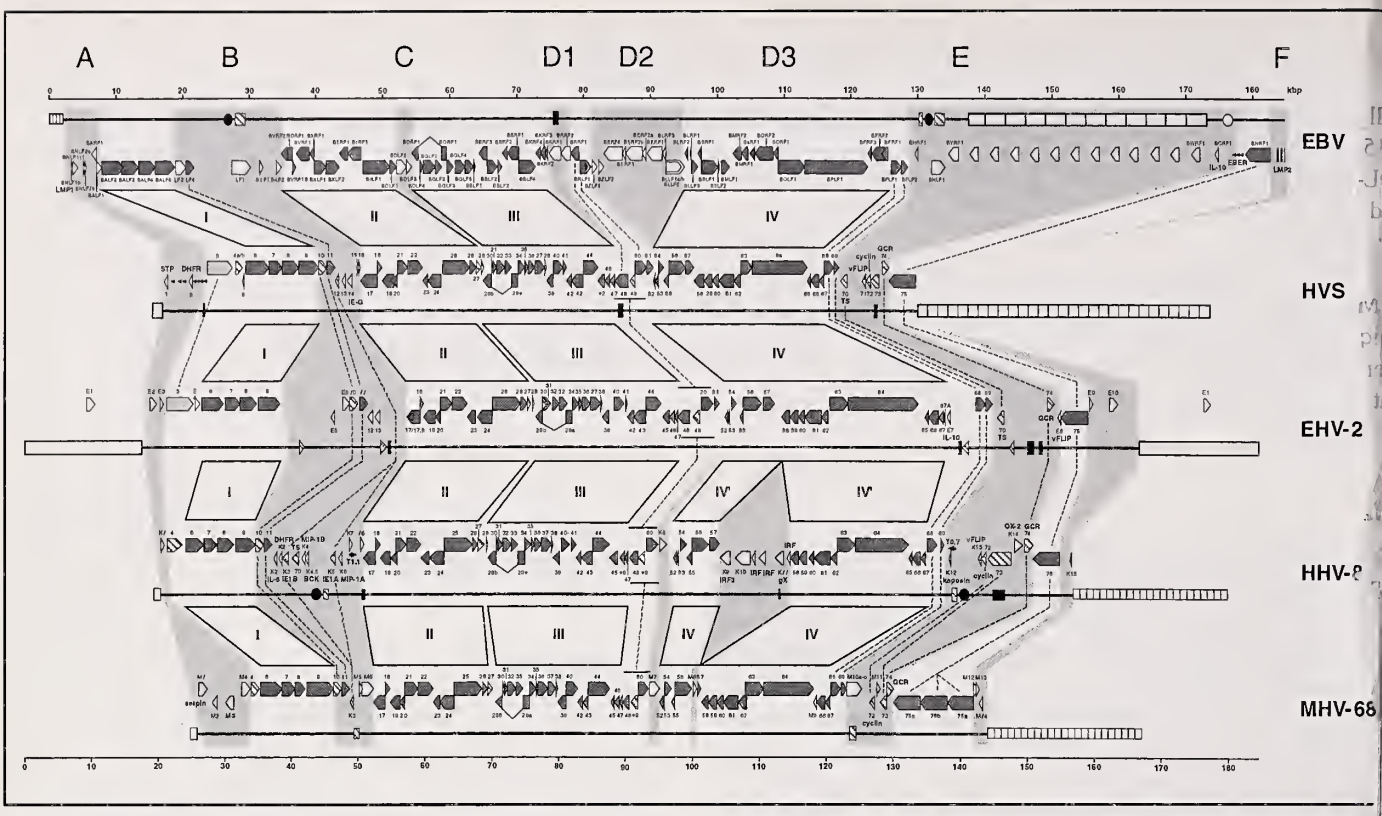


Fig. 1. Comparative organization of the genomes of five gamma herpesviruses. The scale diagram shows the relative size, location, and leftward or rightward orientation of all of the known or predicted genes of each virus based on complete DNA sequence information for human γ1 Epstein-Barr virus [EBV; *(1)*]; the prototype simian γ2 virus of squirrel monkeys, herpesvirus saimiri [HVS; *(2)*]; equine herpesvirus two [EHV-2; *(3)*]; Kaposi's sarcoma-associated herpesvirus or human herpesvirus eight [HHV-8; *(4)*]; and murine herpesvirus one [MHV68; *(5)*]. Genes within conserved blocks labeled I, II, III, and IV are common to all mammalian and avian herpesviruses of the alpha, beta, and gamma classes. Each individual DNA genome is drawn as a solid horizontal line with open boxes representing large terminal and internal direct or tandemly repeated domains. Individual genes with evolutionary homologues in all sequenced gamma herpesviruses are shown as solid bars and those that are prese in only one or several of the five gamma herpesviruses are given as open hatched arrows. The standard orientation of EBV used by Baer et al. *(1)* has bee inverted to match that of the other genomes. Size scales in kilobase pairs a given for EBV at the top and for EHV-2 at the bottom. The major divergent lo (DL-A, B, C, D1, D2, D3, E, and F) are indicated by gray shading and vertic dashed lines denote positions of homologous genes within these loci. Small sol or hatched boxes and arrows depict short GC-rich repeats and inverted pali dromic areas, whereas solid and open circles indicate structurally complex d mains with known or presumed origin-like features, e.g., ori-Lyt(L) or ori-Lyt((●), and ori-P (○) in EBV and the duplicated proposed ori-(L) and ori-(domains in HHV-8 (●).

large direct terminal repeats rather than the otherwise typical multicopy short terminal repeats *(3)*.

The genome organization of EBV differs from that of HVS primarily by the presence of the unique B-cell latency genes and other features, including ori-P, EBNA-1, EBNA-2, EBNA-3abc, LPNA, EBERs, and LMP-1 and 2 as well as the B-cell receptor glycoprotein, the lytic cycle triggering DNA-binding bZip protein ZTA, the large (10 × 3.1 kilobase [kb]) internal IR repeats, and the duplicated ori-Lyt domains (DS-L and DS-R). In their place, HVS has genes encoding the TIP, STP, and vDHFR proteins together with the HSUR small RNAs at the left-hand end (DL-A region) and the vTS(ORF70), vFLIP(ORF71), vCycD(ORF72), ORF73, and an IL8R-like vGCR (ORF74) gene in DL-E. HHV-8 is clearly not a γ1 virus and it has some features that closely resemble HVS, such as the vFLIP (K13) *(7)*, vCycD, LANA(ORF73), and vGCR gene block in DL-E *(8)*, and the presence of short multicopy terminal tandem repeats, but it also differs significantly from HVS by its higher overall GC content (55%), lack of CpG-suppression, insertion of the IRF gene block (DL-D3), and major additions or alterations in several of the divergent loci, especially in DL-A, DL-B, and DL-F. At the present time, equally compelling arguments can be made either for classifying HHV-8 along with all of the other no EBV-like gamma herpesviruses together in a super γ2 subgrou or for giving both HHV-8 and EHV-2 separate status as γ3 an γ4 viruses. Alternatively, it might be wise to abandon any a tempt at rational subdivisions within the gamma herpesvirus

Novel Features of KSHV/HHV-8

Major features of the HHV-8 genome were described prev ously by Moore et al. *(9)*, Cesarman et al. *(8)*, Russo et al. *(4* and Nicholas et al. *(10)*. These include the lack of all EB specific features and the absence of most HVS specific feature except for the presence of vTS and vDHFR genes (both at ne locations) and retention of the vFLIP, vCyc-D, LANA, a vGCR block. Several vIRF-like genes (e.g., ORF-K9) *(11)* a other potential novel genes (ORF-K8, K8.1, and K9–K11) lie the DL-D2 and D3 regions, and a vOX-2 homologue (ORF-K1 has been inserted adjacent to the vIL8R-like vGCR gene DL-E. A series of novel viral-encoded cytokine genes m within DL-B e.g., vIL-6 (K2), vMIP-1A (MIP-I, K6), vMIP-1 (MIP-II, K4), and vBCK (K4.1), together with vTS, vDHF and two LAP/PHD class zinc finger protein genes referred to IE1-A (K5) and IE1-B (K3) *(10–13)*. Two novel abundant sh

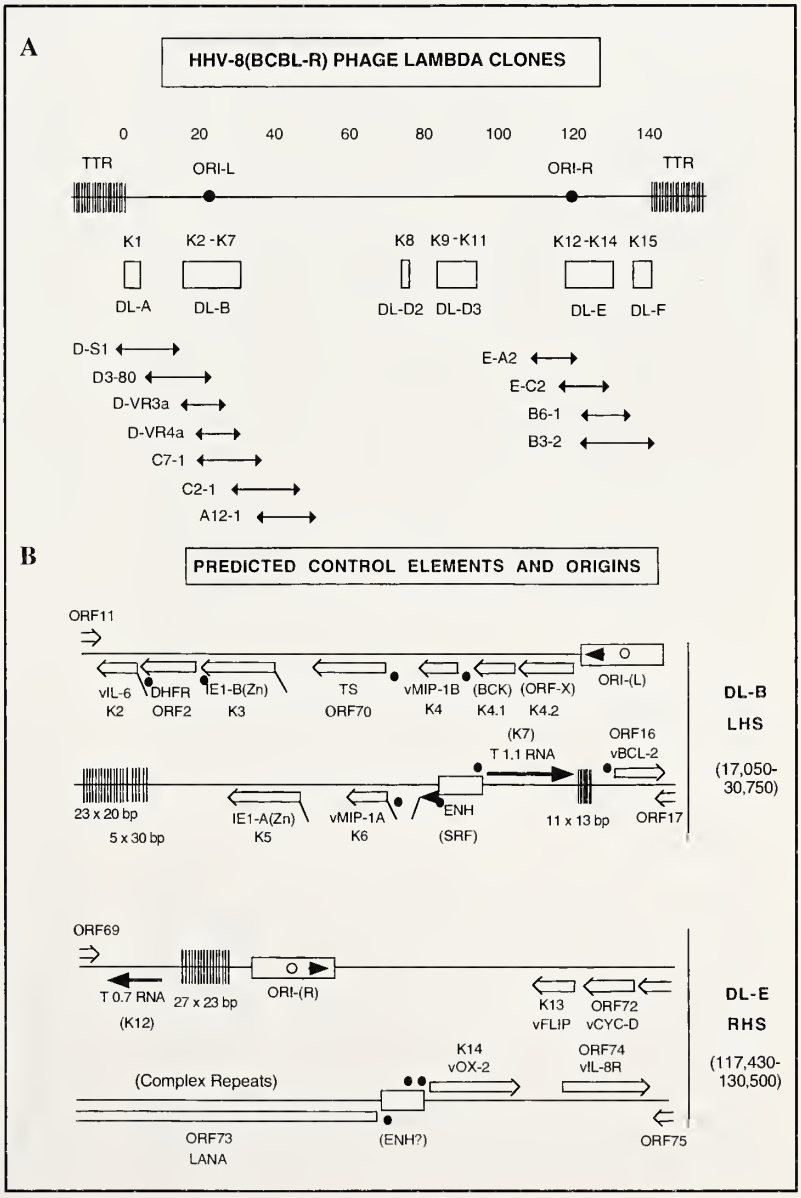
NAs (T1.1/Nut/K7 and T0.7/K12), which are unlike the SJRs of HVS or the EBERs of EBV, have been described *15)* that also map within DL-B and DL-E, respectively. Fi-v, two small novel open reading frames (ORFs), K1 and occur at the left and right ends of the HHV-8 genome A and DL-F) and are positional equivalents of the latency transformation-related LMP-1 and LMP-2 membrane gly-rotein genes of EBV and the strain variable TIP and STP es of HVS *(16,17)*.

Many of these genes appear to be likely to contribute to the ue biology of HHV-8, but evidence for any direct involve-in disease processes must await extensive analysis of the erns of gene expression in different cell types in Kaposi's oma (KS) and other tumors, etc., and detailed examination he functions of their protein products.

Materials and Methods

The sources of the body cavity-based lymphoma (BCBL) and DNA samples and cell lines used have all been described previously, as well as the procedures used for library preparation, subcloning, and genomic or PCR sequencing of HHV-8 DNA *(10,13,18–20)*. Procedures used for transcriptional and functional analysis of vIL-6 were also described previously *(13)*. Details of the functional analysis of vTS and vDHFR, the transcriptional evaluation of the DL-B region, and the sequence data for the LHS and RHS ends of HHV-8(BCBL-R) will be presented elsewhere.

New HHV-8 phage lambda clones not previously described include λD-S1 (–1.8/14.5-kb), λE-A2 (108.3/120.5-kb), and λE-C2 (115.7/129.0-kb), all in the λDASHII background, and λB3–2 (120.6/137.4) in the λEMBL3 background (Fig. 2, A). These were all isolated from either of two phage libraries generated from size selected Sau3A partial digests of a BCBL tumor DNA sample described previously that we refer to as BCBL-R *(10,19)*. λD-S1 was identified by hybridization to a probe representing the right-hand terminus of λD3–80(ORF6), λE-A2 by hybridization to a T0.7 gene probe generated by polymerase chain reaction (PCR) based on sequence data from Zhong et al.



. 2. (A) Diagram illustrating the map locations of two sets of rlapping phage lambda clones of HHV-8(BCBL-R) relative to positions of the divergent loci (DL) and other key features of the ome. The six divergent loci containing gamma-2 specific or IV-8 specific genes (K numbered ORFs) are shown by open es. The relative positions (in kbp) of 11 selected phage lambda nes characterized in this study are shown as horizontal arrows. R indicates the 801-bp tandem repeats at the termini *(4)*. (B) uctural organization of the genes and repeated regions in the IV-8 divergent loci DL-B and DL-E relative to predicted pro-ter control elements and origin-like features. The diagram shows liminary transcript maps of the 13-kb divergent transcription main in DL-B (LHS), and the area around the inverted duplica-n of the 1050-bp ori-like domain in DL-E (RHS). Map coordi-tes are given in parentheses. Open arrows show the size and entation of ORFs and solid arrows show the two abundant RNA ecies referred to as T1.1 (lytic) and T0.7 (latent). Multiple adja-nt vertical lines indicate four sets of short GC-rich tandem repeats d solid circles represent recognized promoter domains containing sely packed TATAAA-like, ATF, SP-1, AP-1, CTF, or ZRE otifs. The proposed divergent enhancer (core SRE motifs in DL- and the origin-like domains are shown as open boxes. Open cles indicate the core origin $(TA)_n$ motifs and solid arrowheads dicate three copies of an approx 90-bp region, one representing e leftwards upstream promoter and leader associated with the L-B ENH domain, and the others are partially conserved versions the same sequence included in the ori domain duplications.

(14), λE-C2 by hybridization to a probe generated from the right-hand terminus of λB6–1(ORF71), and λB3–2 by hybridization to a probe representing the left-hand terminus of λB6–1 (ORF75). The map coordinates of those and our previously described BCBL-R phage lambda clones *(10,19)* were confirmed by restriction enzyme cleavage mapping and terminal sequencing to correspond to genome positions in the HHV8(BC-1) genomic DNA sequence of Russo et al. *(4)* as defined in parenthesis as follows: λD-VR3A (14.7/26.7 kb) and λD-VR4A (18.5/31.2 kb) in the λDASHII background and λD3–80 (5.4/22.9 kb), λC7–1 (19.3/37.9 kb), λC2–1 (27.8/47.2 kb), λA12–1 (35/51 kb), and λB6–1 (121.7/134.4-kb), all in the λEMBL3 background. New plasmid subclones used for additional sequencing included pDJA60 (*Bam*HI/*Eco*RI 2.9 kb encompassing ori-L), pDJA61 (*Bam*HI/*Bam*HI, 3.5 kb encompassing the LHS TTR, ORF-K1A, and part of ORF4), pDJA62 (*Hin*dIII/*Sal*I, 3.3 kb encompassing ORF-K15A and the RHS TTR), plus pDJA63 (*Eco*RI/*Eco*RI, 6.0 kb encompassing T0.7 and ori-R).

Results

Extension of the BCBL-R Phage Library

We have described previously the identification and sequencing of nine novel captured cellular genes mapping between ORF11 and ORF17 within DL-B on the LHS of the HHV-8(BCBL-R) genome, as well as vCycD (ORF72) and an IL8R-like vGCR(ORF74) within DL-E on the RHS *(10,13,19–21)*. In the course of that work, two sets of contiguous phage lambda clones encompassing ORF6 to ORF31 (now defined as map coordinates 5.4–51 kb) and ORF72 to ORF75 (map coordinates 120.6–134.4 kb) were assembled. These two blocks have subsequently been extended both internally between ORF64 and ORF72 to completely cover the remainder of the DL-E region as well as toward the ends of the genome to encompass DL-A at the extreme LHS boundary and DL-F at the extreme RHS boundary, including proximal copies of the terminal tandem repeats (TTR). New phage clones λD-S1, λE-A2, λE-C2, and λB3–2 were identified by plaque hybridization with probes representing the left-hand side (LHS) end of λD3–80 *(10)*, the T0.7 transcript *(14)*, and the right-hand side (RHS) end of λB6–1 *(19)*, and short regions at the termini of each insert were sequenced. A summary of the map locations of a total of seven characterized phage lambda clones covering 53 kb at the left end (positions −1.8 to 51 kb), and four other selected lambda clones covering 30 kb at the right end (positions 108 to 138 kb) of the HHV-8(BCBL-R) genome are shown in Fig. 2, A.

An Inverted Duplication Contains Potential Origin Features

Our analysis has revealed the presence of a 1050-bp domain in DL-B between ORF-K4.2 and 1E1-A (K5) that has typical features of a herpesvirus lytic cycle DNA replication origin, including clustered consensus binding motifs for AP-1 (JUN/FOS) and CTF, multiple short repetitive motifs containing *Xca*I and *Pvu*II sites that resemble features of EBV ori-Lyt, two *Fsp*I/*Sph*I motifs matching sites found within CMV ori-Lyt, and two long AT-rich palindromes, including ATATATATATATATATAT, which resemble the core loop structures of HSV ori-L and ori-S and are also found in the HHV6 and HHV7 lytic origins. Most remarkably, the entire 1050 bp is duplicat as an almost identical inverted copy between T0.7 (K12) a vFLIP (ORF71) toward the other end of the genome with DL-E. Notably, the positions of the two putative ori-Lyt d mains in HHV-8 closely match those of the two directly repeat 800-bp ori-Lyt domains in EBV and, as in EBV, both also directly adjacent to high GC-content tandem repeats. In HHV- these repeats are represented as both 20-bp and 30-bp tande arrays in the LHS region (DL-B) and as two types of 23- tandem repeats in the RHS region (DL-E), and they are po tionally analogous to the 102-bp *Pst*I and 125-bp *Not*I tande repeats in EBV, respectively. Interestingly, MHV68 also h equivalent sets of 40-bp and 100-bp GC-rich tandem repeats these locations, and EHV-2 has a total of four sets of both dire and inverted duplicated palindromic regions with $(AT)_n$ c loops and multiple AP-1 binding sites on the stems within adjacent to the DL-B and DL-E regions. Only HVS appears lack these features. The locations of the predicted LHS ori-L domains within DL-B(ori-L) and DL-E (ori-R) are shown in F 2, B.

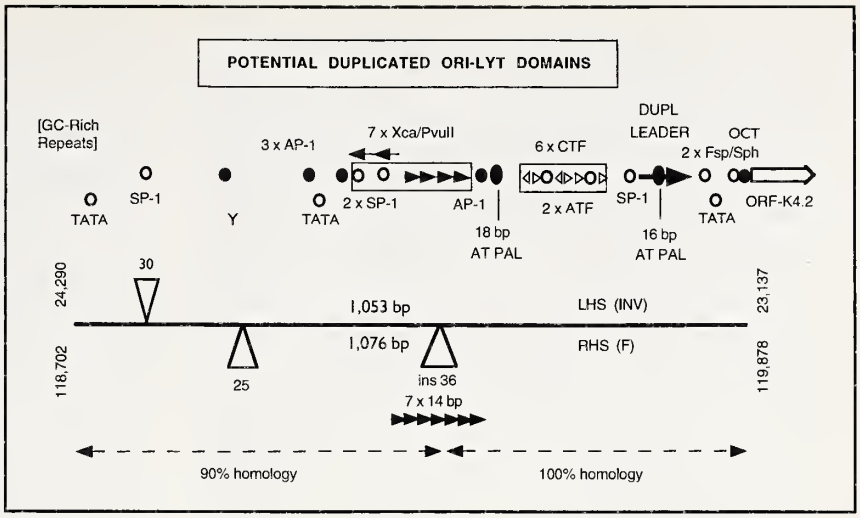
Evaluation of Transcriptional Control Elements

We recognize eight distinct promoter domains within the kb DL-B region, including a 500-bp divergent promot enhancer (ENH) between vMIP-1A and T1.1 that functions a relatively weak constitutive plus butyrate and TPA-inducil promoter in both orientations in transient reporter gene assays Vero, HeLa, and U937 cells. This domain contains a predic rightwards noncoding leader exon of 45 bp that together with upstream TATA-box region is duplicated in a partially c served manner within the two proposed ori-Lyt domains. In fa within the DL-B region, the arrangement of the leftwar oriented leader (ENH version) together with the vMIP-1A a IE1-A genes, followed by the leader (ori-L version) plus vM 1B and 1E1-B genes in the same orientation, suggests anot ancient duplication that may perhaps have occurred prior to acquisition of the other inserted cellular complementary DN (cDNAs) in DL-B (Fig. 3). Note that the 13-kb DL-E dom also appears to represent a divergent transcriptional unit wit single potential ENH-like domain associated with several r ognizable transcription motifs between ORF73 (LANA) a ORF-K14 (vOX-2), but no functional data are available here yet.

Curiously, both orientations of the divergent ENH-like con domains in DL-B and DL-E contain consensus 7-bp ZRE bi ing motifs *(22,23)* either very close to or slightly downstre from the TATA motifs, and both orientations of the DL-B E proved to be down-regulated by cotransfection with the E ZTA transactivator in Vero cells (Chiou C-J, Hayward GS: c not shown). Many of the proximal promoters also contain ad cent ZRE motifs either upstream or downstream of their TA motifs, and several of the potential ZREs in DL-B are a classical JUN/FOS binding AP-1 sites. These findings raise possibility that some key HHV-8 lytic cycle promoters may regulated negatively by EBV in dually infected BCBL cell li and tumors, and they also reinitiate old unsolved questions ab whether there may be a cellular homologue of the EBV l cycle triggering protein ZTA, which HHV-8 may use.

Expression of messenger RNAs (mRNAs) from most of

. 3. Detailed comparison and structural features of the licated HHV-8 ori-like domains in DL-B (inverted ori-ation) and DL-E (forward orientation). Arrows, boxes, symbols in the upper segment denote various subre-ns and characteristic motifs, and the solid lines in the er segment indicate the positions of significant nonho-logous blocks within the left-hand copy (above the line) npared with the right-hand copy (below the line). The bp insertion in ori-(R) compared with ori-(L) represents t of a 7 × 14-bp set of diverged repeats of the *Xca/Pvu*II tif. To the right of the insertion, the two copies in both IV-8(BC-1) and HHV-8(BCBL-R) are virtually identical r 500 bp, whereas to the left of the insertion the left and ht copies are approximately 95% identical outside the erged blocks near the left end.



nes in DL-B (e.g., vDHFR, vTS, vMIP-1B, IE1-B, IE1-A, and 1.1) has been found to occur only after butyrate induction of e HHV-8 lytic cycle in the HBL-6/BC-1 (EBV-positive) cell e and after either butyrate or TPA induction in the BCBL-1 BV-negative) cell line. In contrast to TPA induction, butyrate duction appears to be abortive (to the extent that the viral RNAs are induced only transiently and disappear after be-een 24 and 48 hours), and curiously TPA does not induce the tic cycle genes efficiently in the dual EBV- and HHV-8-fected HBL-6 cell line *(10,24,25)*. vIL-6 gives low-level con-tutive expression in the HBL6/BC1 cell line but is also greatly creased by induction *(11,13)*. Both the predominant vIL-6 .0-kb) and vMIP-1B (0.8-kb) mRNAs produced after induc-n have been found by primer extension analysis to represent ort monocistronic unspliced mRNA species (Poole L, Ciufo Hayward GS: data not shown). However, both the IE1-A and 1-B genes, which lack identifiable proximal promoters, con-n excellent splice-acceptor motifs, and the vIL-6 and vMIP-A genes also appear to have the potential to occur as either liced or unspliced versions (Fig. 3).

nctional Activity of the Viral IL-6, Dihydrofolate eductase (DHFR), and Thymidylate Synthetase S) Genes

Considering that the multistep captures of those cellular NAs are likely to have been relatively ancient evolutionary ents, most such genes that are retained are expected to be nctionally active, although they may have altered their func-ns in subtle ways either by broadening the substrate or target ecificity or by acquiring constitutive rather than conditional tivity that may lead to interference with normal host mecha-sms or apoptotic or immune responses, etc. We have previ-sly shown that the HHV-8-encoded vIL-6 produced from an V2-vIL-6 plasmid in transfected rat embryo fibroblasts substi-tes for human IL-6 in promoting growth of B9 mouse my-oma cells in a manner that uses the interleukin 6 (IL-6) recep-r *(13)*. It also mimics human IL-6 by inducing specific mRNA d DNA-binding transcription factors associated with the ute-phase response in HEP-3B cells [*(13)*; Wan X.-Y., Nicho-s J: manuscript in preparation)]. Some properties and functions the HHV-8-encoded Cyc-D (ORF72), the IL-8R-like vGCR (ORF74), vMIP-1A (K6), and vIL-6 (K2) from HHV-8(BC1) have also been described by others *(8,11,26,27)*.

We have also more recently cloned the HHV-8-encoded vDHFR and vTS genes into bacterial glutathione *S*-transferase (GST) and mammalian Flag-tagged vector systems (Sarisky R, Ciufo D, Zhang X, Chiou C.-J, Hayward GS: manuscript in preparation). Stable DHFR-negative CHO cell lines constitutively expressing the tagged vDHFR have been generated by Neo co-selection and these proved to be susceptible to cell killing with either 10 or 100 m*M* methotrexate. Furthermore, a methotrexate-resistant high copy number subclone efficiently bound to a fluorescent tagged dihydrofolate substrate as assayed by fluorescence-activated cell sorter analysis. Similarly, a transfected GST/vTS fusion protein plasmid proved both to rescue a thymidylate synthetase-negative *Escherichia coli* strain in Thy^- medium and to make the host cells susceptible to killing by fluorouracil. All of these properties indicate that the vIL-6, vDHFR, and vTS genes retain the normal functional activities of their cellular homologues, although we have yet to discover whether or not their substrate or target specificities have been altered. Evidence that the HHV-8 encoded vBcl-2 (ORF16) protein functions similarly to human Bcl-2 to block apoptosis has also been presented *(20,28)*.

Terminal Heterogeneity on the Right-Hand Side

Selective sequencing of relevant terminal segments of HHV-8(BCBL-R) DNA subcloned into plasmids from the phage lambda clones λD-S1, λE-A2, and λB3–2 has also been carried out (Alcendor D, Zhong J.-C, Wan X, Nicholas J, Guo H, Reitz M, et al: manuscript in preparation). At the extreme RHS end of the unique segment of the HHV-8(BCBL-R) genome a 3.3-kb *Hind*III to *Sal*I insert from λB3–2 subcloned in plasmid pDJA62 was completely sequenced by standard M13 shotgun procedures to extend our previous UPS75 DNA sequence (Genbank accession no U85269). The results revealed a match to the published HHV-8(BC-1) sequence across the N-terminus of ORF75, but with a pattern of increasing nucleotide differences for 400 bp between positions +120 to −280 relative to the initiator codon, followed by 2300 bp of completely different sequence toward the end of the unique region, and then returning to a perfect match (for at least 700 bp) with the 801-bp HHV-8(BC-1) TTR

unit sequence (Fig. 4). The published HHV-8(BC-1) sequence contains 3067 bp of non-TTR sequence upstream of the initiator codon for ORF75 but is incomplete at the junction between the unique and TTR domains because of an unstable (apparently unclonable) G plus A-rich variable repeat region estimated to be 2–3 kb in size *(4)*. In contrast, our HHV-8(BCBL-R) sequence contains 2579 bp between the ORF75 initiator codon and the junction with the first copy of the TTR, which begins at position 652 within the complete 801-bp TTR sequence given by Russo et al. *(4)*. There are no G plus A-repeats at the TTR junction in BCBL-R, but we believe, nevertheless, that this represents the complete undeleted form because PCR amplification with primers that span the junction from the unique region into the RHS TTR gave identical-sized products from the original BCBL-R tumor DNA sample, the λB3–2 phage DNA, and the pDJA62-subcloned plasmid DNA.

On the basis of our previous identification of three distinct subtypes of HHV-8 in the ORF26 and ORF75 gene regions *(18)*, the published sequence for HHV-8(BC-1) *(4)* indicates that it represents a C subtype at the RHS end but an A subtype elsewhere, whereas the HHV-8(BCBL-R) genome is an A subtype throughout. DNA from the RHS of our prototype C strain proved to be identical to that in both the BC-1 and HBL6 cell lines in two distinct KS lesions from this patient (ASM72/76) for as far as at least 1200-bp upstream from the ORF75 initiator codon. In contrast, all three other A strains tested (C282, AKS1, and BCBL-1) proved to be nearly identical to BCBL-R between ORF75 and the TTR, although our prototype B strain (431KAP) differed by the presence of both a 10-bp insertion and a 55-bp deletion across the unique sequence/TTR boundary. Therefore, whereas the HHV-8 C strains are dramatically different from the A and B strains over at least 2500 bp (and perhaps up to 5000 bp) at the extreme RHS end of the unique region, B strains proved to be only slightly different from the A strains in this region (Fig. 4).

Interestingly, both the A/B and C strain unique regions upstream and to the right of ORF75 have the potential to encode complex multiply spliced leftwards-oriented mRNAs for hydrophobic membrane proteins that have an equivalent position and orientation to that of EBV LMP-2. Attempts to characterize the very different ORF-K15A/B and ORF-K15C transcripts and gene products are in progress.

Terminal Heterogeneity on the Left-Hand Side

We also carried out plasmid and PCR primer-based sequencing over a 1500-bp block from genomic nucleotide positions 40–1540, encompassing the entire ORF-K1 and the C-terminal segment of ORF4 for BCBL-R (pDJA61) and eight other HHV-8 samples from various sources. Although this segment of the conserved ORF4 coding region showed only three nucleotide variations over 400 bp among A, B, and C strains, which is equivalent to the level found for ORF26 and ORF75 (i.e., approximately 1% differences; 18), the 840-bp (289 amino acid [aa]) ORF-K1 coding region proved to display much greater variability (Figs. 4, 5, and 6). Most dramatically, our prototype B strain HHV-8(431KAP) differed by 14.6% (127 positions) at the DNA level and 29% (84 amino acids) at the protein level from the published data for HHV-8(BC-1). Similarly, our prototype C strain HHV-8(ASM72/76) differed by 5.8% (49 positions) at the DNA level and 14% (39 amino acids) at the protein level from HHV-8(BC-1). The B and C strains also differed from each other at 125 nucleotide positions (and 83 amino acids), although again the subtype C DNA genomes present in two distinct lesions from the same patient (ASM72 and ASM76) were identical throughout. The HBL6 and BC-1 cell lines were derived independently from the same patient and both are A subtype sequences throughout (except at the RHS or ORF-K15 region) and as expected they proved to have identical ORF-K1 DNA sequences. Among the other A strains tested, BCBL-R exhibited 19-bp (2.0%) and 13 amino acid (4.5%) changes, BCBL-1 gave 27-bp (2.8%) and 22 aa (8%) changes, and C282 gave 16-bp (1.7%) and 11 aa (4%) changes relative to HHV-8(BC-1). However, the ORF-K1 regions in BCBL-R, C282, AKS1, AKS2, and AKS4 differed by only 2–5 bp relative to one another and all are considered to be members of the A1 substrain or clade. Even the junctions between the LHS unique sequence and the TTR differed significantly between BCBL-R(A1), BCBL-1(A3), and BC1(A2), but no data about the LHS junc

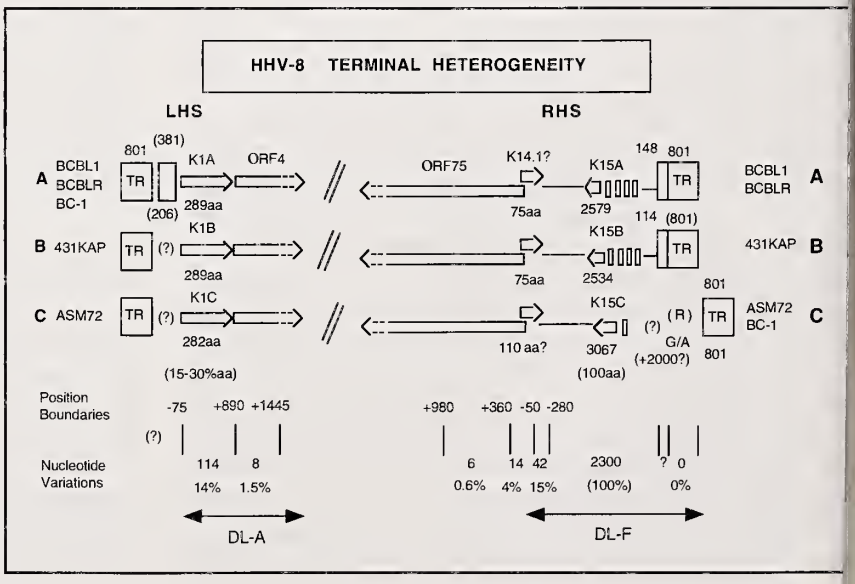
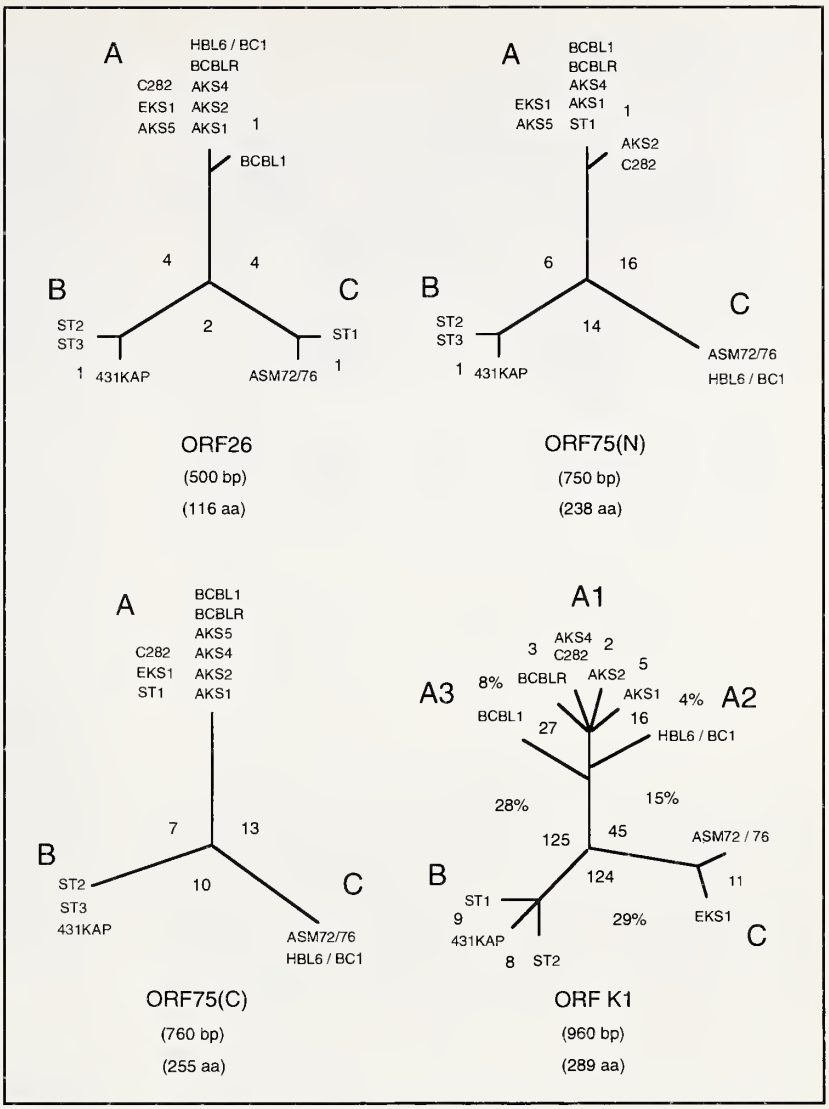


Fig. 4. Locations of major heterogeneous domains in the DL-A and DL-F regions at the extreme left and right genomic termini of the A, B, and C Substrains of HHV-8. ORFs are depicted as open arrows and the most proximal complete and partial (where known) terminal repeat units are indicated by large open boxes. Multiple potential exons in ORF-K15 are shown as small open boxes. Differences in organization and coding content of strains A, B, and C at the termini include the following: 1) the LHS junctions of the TR and unique regions differ even between the A1, A2, and A3 substrains; 2) the protein coding content of ORF-K1A and ORF-K1C differ by 15% at the amino acid identity level, whereas ORF-K1B differs from both of the other subtypes by nearly 30%. ORF-K1C genes also have a consistent five amino acid deletion; 3) a potential short rightwards ORF-K14.1 coding region differs in size and coding content in the C strains compared with A and B strains; 4) the entire unique 2.5- to 5-kb region between ORF75 and the TTRs, which encompasses the spliced ORF-K15 genes, is different in some C strains compared with A and B strains; and 5) the predicted splicing patterns of ORF-K15 and the positions of the unique/TTR junctions on the RHS differ between A and B strains.



ig. 5. Summary of the levels of nucleotide differences ob-rved between sequenced HHV-8 A, B, and C strains within ur distinct coding regions of the genome. The first three lo-ations, namely, ORF26 and two adjacent blocks within ORF75 ferred to here as ORF75(N) and ORF75(C) *(18)*, display typi-l 0.8%–1.5% nucleotide variations between the A, B, and C rains with 10-fold less among isolates from within the same rain designation. However, the ORF-K1 region displays 10- to)-fold greater variation, with C strains differing by 15% at the NA level and up to 30% at the protein level from A and B rains, and with C strains differing from A strains by up to 8% the DNA level and 15% at the protein level. Even within the strains, distinct A1, A2, and A3 ORF-K1 substrain patterns an be discerned, with differences of 2%–3% at the DNA level nd 4%–8% at the protein level between them, but falling to etween 0.2% and 0.5% nucleotide differences at the DNA evel within the A1 subgroup. Individual patient samples tested epresent the following: U.S. AIDS-associated BCBL tumors BC1/HBL6, BCBL-1, and BCBL-R); African classical KS 431 KAP) and African AIDS-associated KS (ST1, ST2, and T3), U.S. classical KS (EKS1); East Coast U.S. AIDS-ssociated KS (C282, AKS1, AKS2, AKS4, and ASM72/76). Iote that three of the 13 genomes shown (namely, BC-1/HBL6, T1, and EKS1) represent chimeras with the structures A/A/C, /C/A, and C/A/A, respectively, within their LHS (ORF-K1), entral (ORF26), and RHS (ORF75N and ORF75C) blocks.

ions are available as yet for the prototype B and C strains (Fig. 4). Summary diagrams comparing these ORF-K1 variation lev-ls with those in the three previously analyzed more conserved egions in ORF26, ORF75(N), and ORF75(C) *(18)* are presented n Fig. 5.

The potential 289 aa membrane protein encoded by the pre-licted unspliced ORF-K1 gene is a positional and orientation counterpart to LMP-1 of EBV (Fig. 1), but it lies in an inverted orientation relative to the highly variable STP (and TIP) trans-forming proteins of HVS, and it does not closely resemble any of these known gamma herpesvirus transformation related pro-teins in structure or amino acid content. The fact that all versions tested remain in frame, despite the presence of both a 15- and a 5-bp deletion in the prototype C strain, attests to the presumed functional importance of the ORF-K1 gene product. Interest-ingly, the major features of 12 Cys residues and 8 N-glycosylation sites within the N-terminal 175 amino acids of ORF-K1 are nearly fully conserved and hydrophobic blocks be-tween positions 1–22 and positions 227 and 251, which are not altered significantly by the strain variations, probably serve as a signal peptide and a membrane anchoring transmembrane do-main, respectively (Fig. 6, A). Therefore, the bulk of the ORF-K1 protein has the appearance of an extracellular receptor that displays a similar level of variation to that encountered in the HIV ENV protein. In contrast, the smaller C-terminal cytoplas-mic domain is identical between most A and C strains, but differs by 30% at the amino acid level in the B strains. An illustration of the extent of the amino acid changes over the two most variable extracellular domains of the ORF-K1 protein be-tween amino acids 52–92 and between 191 and 228 is shown in Fig. 6, B.

Discussion and Conclusions

Although the recently characterized herpesvirus KSHV or HHV-8 *(6)* has not yet been demonstrated directly to be trans-missible from tumor cells to other target cells or from patient to patient, most of the epidemiologic, serologic, and molecular bi-ology evidence currently available suggests that it is likely to be an essential infectious causative agent of both classical and AIDS-associated KS *(29–34)*. Herpesvirus particles can be in-duced from cultured HHV-8-positive EBV-negative BCBL cell lines *(25)* and both the lytic and latent HHV-8-encoded abundant RNAs (T1.1 and T0.7) can be detected in spindle cells and mononuclear cells in KS lesions *(8,14,35,36)*. The genome of

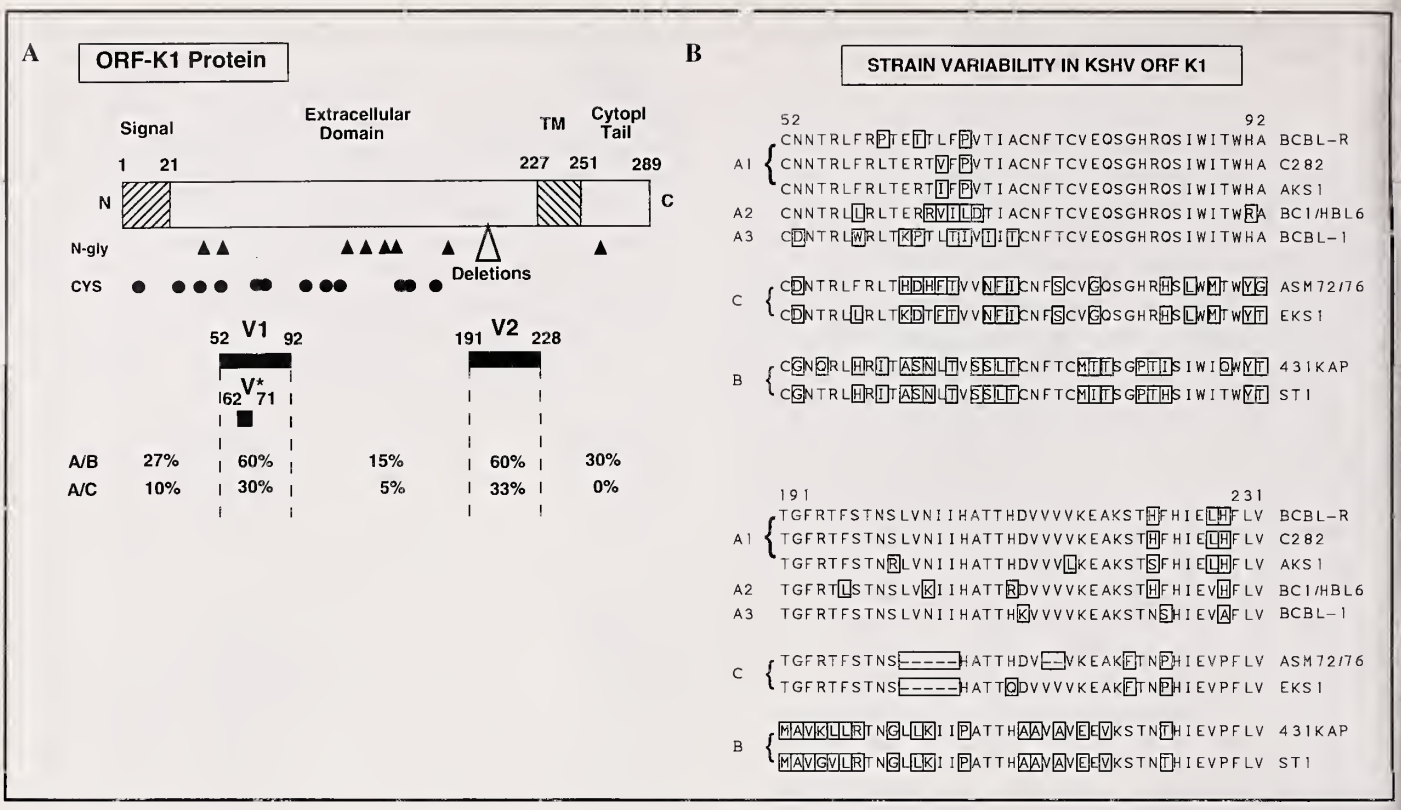


Fig. 6. Organization of the HHV-8 ORF-K1 protein and illustration of the level of amino acid divergence observed within and between the A, B, and C strains. **A)** Domain structure diagram of the predicted 289 amino acid ORF-K1 protein, illustrating the size and locations of the N-terminal signal peptide and C-terminal transmembrane domain and cytoplasmic tail. The positions of N glycosylation sites (NXS/T) and fully conserved Cys residues in the extracellular domain are denoted by solid triangles and circles, respectively. Overall amino acid identity values (%) between the prototype A(C282), B(431KAP), and C(ASM72) genomes are given for the two defined major variable domains (V1 and V2). V* denotes the hypervariable region between amino acids 62–71. **B)** Amino acid sequences of two 41-residue blocks representing the major variable regions (V1 and V2) are shown for nine distinct patient samples, together with isolate names and substrain designations. Boxed amino acid symbols represent deviations from the commonest residues found at each position and dashes represent deleted residues.

the virus is essentially intact and probably circular in the BCBL cell lines, but these points are less clear for the KS tumor versions *(9,36,37)*.

Whether virus-infected cells only provide cytokine support for the other cells in KS lesions or are themselves capable of proliferating is also the subject of considerable debate at present. However, the presence, expression, and functional integrity of the many captured cellular cDNA homologues identified enhance the plausibility that this virus, like EBV and HVS, probably has the potential to directly influence the proliferation state of its host cells whilst in a latent or nonpermissive mode. Expression of the viral-encoded MIP-1-like inflammatory chemokines (even if it occurs only during lytic cycle infection) implies that infected cells may be able to profoundly alter neighboring cells, even to the extent of influencing patterns of HIV infectivity through blocking of the CCR5 co-receptor *(11,13,39,40)*. Furthermore, it has long been known that IL-6 and related cytokines play a major role in the biology of KS spindle cells *(41–43)*, as well as in multiple myeloma *(44,45)* and MCD *(46)*, and recent evidence by *in situ* hybridization (Cannon J, Ambinder R: personal communication) and immunohistochemistry *(11)* that vIL-6 is expressed *in vivo* in most BCBL tumor cells greatly increases the likelihood that such scenarios may be valid.

Currently, first generation serologic assays for the level of seropositivity to HHV-8 antigens have indicated that the virus may not be widespread outside the AIDS epidemic, except in isolated pockets in areas associated with classical and endemic KS, including some Mediterranean and middle east countries, such as Italy and Greece, and in equatorial Africa, including Uganda and Zaire *(24,26,30,31,35,47)*. Such a pattern would be very atypical for a well-established herpesvirus, whereas one usually expeets a ubiquitous distribution of a highly infectious agent with nearly universal asymptomatic primary infection at a very young age. Therefore, many intriguing questions about the origins and mode of transmission of the virus arise. If infection with the virus in humans was actually an ancient and widespread event, then numerous distinct isolates or strains would be expected compared with just a single narrowly diverged genome pattern if it had been recently acquired from an exogenous primate source. Similarly, the patterns of genome heterogeneity observed might be expected to reflect whether the virus in patients with AIDS represents a new horizontally transmitted agent, rather than multiple reactivated endogenous latent infections.

Our evidence clearly shows that there are at least three major subdivisions of HHV-8 genomes present in human populations (all three of which are represented in patients with AIDS) and that (with few exceptions) each patient isolate studied to date can be distinguished from that of other patients. Importantly, no differences have yet been found between the genomes present in multiple lesions from the same patient *(10)* or in subcloned

BCBL cell lines with different passaging history from the same tumor (i.e., HBL6 and BC1). Our collection of genomes that were tested came from a wide variety of sources, but interpretation and extrapolation of the current strain variability data are greatly complicated by large differences in the rates of evolutionary divergence in some genes (e.g., ORF-K1) compared with the others and the existence of two distinct alternative RHS ends. The possible presence of chimeric or recombinant genomes containing gene blocks derived from more than one strain lineage is also implied by mismatches between the ORF-K1 subtype and the constant region (ORF26/ORF75) subtypes in some genomes. However, it is not clear as yet whether the low level of variability in the constant region genes (especially ORF26) is sufficient to provide a valid measure of subtype category (Fig. 5). Remarkably, the pattern and degree of amino acid variability in the HHV-8 ORF-K1 protein quite closely resemble those seen in both the HIV ENV gene and in immunoglobulin variable regions, although obviously in this case the source of the variability and the nature of the selective pressures involved are at present a complete mystery. Recent direct PCR screening of HHV-8-positive samples from 22 different U.S. patients has revealed only five with the C-type of RHS compared with 17 of the A type. In contrast, eight were of the C type in the ORF-K1 region on the LHS and 14 were A type, but only three genomes were found to be C type at both the LHS and RHS ends. Surprisingly, no B-strain types have been found as yet amongst either KS or BCBL samples from U.S. patients.

Overall, the results strongly imply that HHV-8 is an ancient human virus with variability and subtype patterns that closely resemble those seen in both EBV and HVS. Whether the A, B, and C strain groupings represent ancient human population bottlenecks, current biologic niche or disease associated selection, or current (pre-AIDS) geographic isolation remains to be determined. However, the fact that five of six KS samples tested from Africa have B strain ORF-K1 regions and that four of our five New York/Baltimore/Washington area AIDS KS samples are closely related A1 strains, whereas the fifth, which was derived from a classic non-AIDS-associated case (EKS1), had a C strain ORF-K1 gene, suggests that we might be able to define clades of the virus similar to the situation in HIV. Obviously, the surprisingly high genetic divergence identified in ORF-K1 among HHV-8 genomes will now provide opportunities to address these questions of origins, geographic preferences, transmissibility, and disease association in considerable detail.

References

(1) Baer R, Bankier AT, Biggin MD, Deininger PJ, Farrell PJ, Gibson TJ, et al. DNA sequence and expression of the B95–8 Epstein-Barr virus genome. Nature 1984;310:207–11.

(2) Albrecht JC, Nicholas J, Biller D, Cameron KR, Biesinger B, Newman C, et al. Primary structure of the herpesvirus saimiri genome. J Virol 1992; 66:5047–58.

(3) Telford EA, Watson MS, Aird HC, Perry J, Davison AJ. The DNA sequence of equine herpesvirus 2. J Mol Biol 1005;249:520–8.

(4) Russo JJ, Bohenzky RA, Chien MC, Chen J, Yan M, Maddalena D, et al. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). Proc Natl Acad Sci U S A 1996;93:14862–7.

(5) Virgin HW, Latreille P, Wamsley K, Hallsworth K, Week KE, Canto AJ, et al. Complete sequence and genomic analysis of murine gammaherpesvirus 68. J Virol 1997;71:5894–904.

(6) Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 1994;266:1865–9.

(7) Thome M, Schneider P, Hofman K, Fickenscher H, Melnl E, Neipel F, et al. Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. Nature 1997;386:517–21.

(8) Cesarman E, Nador RG, Bai F, Bohenzky A, Russo JJ, Moore PS, et al. Kaposi's sarcoma-associated herpesvirus contains G protein-coupled receptor and cyclin D homologs which are expressed in Kaposi's sarcoma and malignant lymphoma. J Virol 1996;70:8218–23.

(9) Moore PS, Gao SJ, Dominguez G, Cesarman E, Lungu O, Knowles D, et al. Primary characterization of a herpesvirus agent associated with Kaposi's sarcoma. J Virol 1996;70:549–58.

(10) Nicholas J, Ruvolo V, Zong JC, Ciufo D, Guo HG, Reitz MS, et al. A single 13 kilobase divergent locus in Kaposi sarcoma associated herpesvirus (HHV8) contains at least nine genes encoding homologues of cellular proteins. J Virol 1997;71:1963–74.

(11) Moore PS, Boshoff C, Weiss RA, Chang Y. Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV. Science 1996; 274:1739–44.

(12) Neipel F, Albrecht JC, Ensser A, Huang YQ, Li JJ, Friedman-Kien AE, et al. Human herpesvirus 8 encodes a homolog of interleukin-6. J Virol 1997; 71:839–42.

(13) Nicholas J, Ruvolo VR, Burns WH, Sandford G, Wan X, Ciufo D, et al. Kaposi's sarcoma-associated human herpesvirus-8 encodes homologues of macrophage inflammatory protein-1 and interleukin-6. Nature Med 1997; 3:287–92.

(14) Zhong W, Wang H, Herndier B, Ganem D. Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma. Proc Natl Acad Sci U S A 1996;93:6641–6.

(15) Sun R, Lin SF, Gradoville L, Miller G. Polyadenylated nuclear RNA encoded by Kaposi sarcoma-associated herpesvirus. Proc Natl Acad Sci U S A 1996;93:11883–8.

(16) Jung JU, Trimble JJ, King NW, Beisinger B, Fleckenstein BW, Desrosiers RC. Identification of transforming genes of subgroup A and C strains of herpesvirus saimiri. Proc Natl Acad Sci U S A 1991;88:7051–5.

(17) Biesinger B, Tsygankov AY, Fickenscher H, Emmrich F, Fleckenstein B, Bolen JB, et al. The product of the *Herpesvirus saimiri* open reading frame 1 (tip) interacts with T cell-specific kinase $p56^{lck}$ in transformed cells. J Biol Chem 1995;270:4729–34.

(18) Zong JC, Metroka C, Reitz MS, Nicholas J, Hayward GS. Strain variability amongst Kaposi sarcoma associated herpesvirus (HHV-8) genomes: evidence that a large cohort of U.S.A. AIDS patients may have been infected by a single common isolate. J Virol 1997;71:2505–11.

(19) Guo HG, Browning P, Nicholas J, Hayward GS, Jiang YW, Sadowska M, et al. Characterization of a chemokine receptor-related gene in human herpesvirus 8 and its expression in Kaposi's sarcoma. Virology 1997;228: 371–8.

(20) Cheng EH, Nicholas J, Bellows DS, Hayward GS, Guo HG, Reitz MS, et al. A Bcl-2 homolog encoded by Kaposi sarcoma-associated virus, human herpesvirus 8, inhibits apoptosis but does not heterodimerize with Bax or Bak. Proc Natl Acad Sci U S A 1997;94:690–4.

(21) Davis MA, Sturzl M, Blasig C, Schreier A, Guo HG, Reitz M, et al. Expression of human herpesvirus 8-encoded cyclin D in Kaposi's sarcoma spindle cells. J Natl Cancer Inst 1997;89:1868–74.

(22) Chang YN, Dong D, Hayward GS, Hayward SD. The EBV Zta transactivator: A member of the B/Zip family with unique DNA-binding specificity and a dimerization domain that lacks the characteristic heptad leucine zipper motif. J Virol 1990;64:3358–69.

(23) Lieberman PM, Hardwick JM, Sample J, Hayward GS, Hayward SD. The Zta transactivator involved in induction of lytic cycle gene expression in Epstein-Barr virus-infected lymphocytes binds to both AP-1 and ZRE sites in target promoter and enhancer regions. J Virol 1990;64:1143–55.

(24) Miller G, Heston L, Grogan E, Gradoville L, Rigsby M, Sun R, et al. Selective switch between latency and lytic replication of Kaposi's sarcoma herpesvirus and Epstein-Barr virus in dually infected body cavity lymphoma cells. J Virol 1997;71:314–24.

(25) Renne R, Zhong W, Herndier B, McGrath M, Abbey N, Kedes D, et al. Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. Nature Med 1996;2:342–6.

(26) Chang Y, Ziegler J, Wabinga H, Katagole-Mbidde E, Boshoff C, Schulz T, et al. Kaposi's sarcoma-associated herpesvirus and Kaposi's sarcoma in Africa. Arch Intern Med 1996;156:202–4.

(27) Arvanitakis L, Geras-Raaka E, Varma A, Gershengorn MC, Cesarman E. Human herpesvirus KSHV encodes a constitutively active G-protein-coupled receptor linked to cell proliferation. Nature 1997;385:347–50.

(28) Sarid R, Sato T, Bohenzky A, Russo JJ, Chang Y. Kaposi's sarcoma-associated herpesvirus encodes a functional Bcl-2 homologue. Nature Med 1997;3:293–8.

(29) Ambroziak JA, Blackbourn DJ, Herndier BG, Glogan RC, Gullett JH,

McDonald AR, et al. Herpes-like sequences in HIV-infected and uninfected Kaposi's sarcoma patients. Science 1995;268:582–3.
(30) Gao SJ, Kingsley L, Li M, Zheng W, Parravicini C, Ziegler J, et al. KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. Nature Med 1996;2:925–8.
(31) Gao SJ, Kingsley L, Hoover DR, Spira TJ, Rinaldo CR, Saah A, et al. Seroconversion to antibodies against Kaposi's sarcoma-associated herpesvirus-related latent nuclear antigens before the development of Kaposi's sarcoma. N Engl J Med 1996;335:233–41.
(32) Dupin N, Grandadam M, Calvez V, Gorin I, Aubin JT, Havard S, et al. Herpesvirus-like DNA sequences in patients with Mediterranean Kaposi's sarcoma. Lancet 1995;345:761–2.
(33) Beral V, Peterman TA, Berkelman RC, Jaffe HW. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? Lancet 1990; 335:123–8.
(34) Boshoff C, Whitby D, Hatziioannou T, Fisher C, Walt J, Hatzakis A, et al. Kaposi's sarcoma-associated herpesvirus in HIV-negative Kaposi's sarcoma. Lancet 1995;345:1043–4.
(35) Boshoff C, Schulz TF, Kennedy MM, Graham AK, Fisher C, Thomas A, et al. Kaposi's sarcoma-associated herpesvirus infects endothelial and spindle cells. Nat Med 1995;1:1274–8.
(36) Staskus KA, Zhong W, Gebhard K, Herndier B, Wang H, Renne R, et al. Kaposi's sarcoma-associated herpesvirus gene expression in endothelial (spindle) tumor cells. J Virol 1997;71:715–9.
(37) Decker LL, Shankar P, Khan G, Freeman RB, Dezube BJ, Lieberman J, et al. The Kaposi sarcoma-associated herpesvirus (KSHV) is present as an intact latent genome in KS tissue but replicates in the peripheral blood mononuclear cells of KS patients. J Exp Med 1996;184:283–8.
(38) Renne R, Lagunoff M, Zhong W, Ganem D. The size and conformation of Kaposi's sarcoma-associated herpesvirus (humanherpesvirus 8) DNA in infected cells and virions. J Virol 1996;70:8151–4.
(39) Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, Lusso P. Identification of RANTES, MIP-1α, and MIP-1β as the major HIV-suppressive factors produced by CD8+ T cells. Science 1995;270:1811–5.
(40) Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima A , et al. HIV-1 entry into $CD4^+$ cells is mediated by the chemokine receptor CC-CKR-5. Nature 1996;381:667–73.
(41) Bailer RT, Lazo A, Ng-Bautista CL, Hout BL, Ness GM, Hegtvedt AK, et al. Comparison of constitutive cytokine release in high and low histologic grade AIDS-related Kaposi's sarcoma cell strains and in sera from HIV+/KS+ and HIV+/KS− patients. J Inter Cytok Res 1995;5:473–83.
(42) Cai J, Gill PS, Masood R, Chandrasoma P, Jung B, Law RE, et al. Oncostatin-M is an autocrine growth factor in Kaposi's sarcoma. Am J Pathol 1994;145:74–9.
(43) Miles SA. Pathogenesis of HIV-related Kaposi's sarcoma. Curr Opin Oncol 1994;6:497–502.
(44) Kawano M, Hirano T, Matsuda T, Taga T, Hori Y, Iwato K, et al. Autocrine generation and requirement of BSF-2/IL-6 for human myelomas. Nature 1988;332:83–4.
(45) Zhang XG, Klein B, Bataille RR. Interleukin-6 is a potent myeloma cell growth factor in patients with aggressive multiple myeloma. Blood 1989; 74:11–3.
(46) Soulier J, Grollet L, Oksenhendler E, Cacoub P, Cazals-Hatem D, Babinet P, et al. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. Blood 1995;86:1276–80.
(47) Whitby D, Howard MR, Tenant-Flowers M, Brink NS, Copas A, Boshoff C, et al. Detection of Kaposi sarcoma associated herpesvirus in peripheral blood of HIV-infected individuals and progression to Kaposi's sarcoma. Lancet 1995;346:799–802.

Notes

Supported by Public Health Service grants CA73585 (G. S. Hayward) and CA06973 (G. S. Hayward and J. Nicholas) from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services. L. J. Poole is a member of the Biochemistry, Cellular and Molecular Biology training program at The Johns Hopkins University School of Medicine (T32 GM07445). R. T. Sarisky and J. Nicholas were supported in part by a Postdoctoral Fellowship and a Junior Faculty Award, respectively, from the American Cancer Society.

We thank Sarah Heaggans for her assistance in the preparation of the manuscript and diagrams and Skip Virgin and Sam Speck for providing MHV68 sequence data prior to publication.

Immunotherapy for Epstein-Barr Virus-Associated Cancers

*Cliona M. Rooney, Marie A. Roskrow, Colton A. Smith, Malcolm K. Brenner, Helen E. Heslop**

Epstein-Barr virus (EBV)-associated lymphoproliferative disease (EBV–LPD) is a frequently fatal complication of organ transplantation and human immunodeficiency virus (HIV) infection. We have studied the safety and efficacy of adoptively transferred, gene-marked virus-specific cytotoxic T lymphocytes (CTLs) as prophylaxis and treatment of EBV–LPD in recipients of T-cell-depleted allogeneic bone marrow. In 42 patients treated prophylactically, no toxicity was experienced. None of these patients developed EBV–LPD, in contrast with eight of 53 (15%) patients who did not receive prophylactic CTL. Three patients who had not received CTL developed aggressive disease and received CTL as treatment. Gene-marked CTL homed to tumor sites and selective accumulation of marker gene was detected in tumor tissues. Tumors regressed completely in two patients, but the third died of respiratory failure. Infused CTLs persisted for up to 3 years *in vivo*, they rapidly reconstituted EBV-specific immune responses to levels seen in normal individuals, and they reduced high viral titers by two to three logs. We are now using autologous EBV-specific CTL to treat patients with relapsed EBV-positive Hodgkin's disease and we are developing methods for the generation of antigen-specific lines. This approach could be applied to patients with HIV who develop EBV–LPD, using CTL derived early in the course of HIV infection. [Monogr Natl Cancer Inst 1998;23: 89–93]

More than 90% of individuals are infected with Epstein-Barr virus (EBV). Primary infection occurs through the oropharynx, where the virus replieates in epithelial cells and infects circulating B lymphocytes *(1,2)*. The virus then persists for the life of the infected individual, in epithelial cells and B lymphocytes. EBV can efficiently transform B lymphocytes *in vitro* into permanently growing lymphoblastoid cell lines (LCLs) if immune T cells are first removed or inactivated *(1)*. Similarly, in the absence of T-cell function *in vivo*, EBV may transform B cells causing lymphoproliferative disease (EBV–LPD) or lymphoma *(3)*. Recipients of stem cells from human leukocyte antigen (HLA)-mismatched family members or unrelated donors are particularly at risk of developing this complication if T cells are removed from the allograft to prevent graft-versus-host disease (GvHD) *(4,5)*.

Unlike disease caused by herpesviruses, such as CMV and HSV, EBV–LPD is associated with the latent phase of the virus. Therefore, drugs such as acyclovir, which target the viral DNA polymerase, are ineffective in prevention or treatment. The nine latency-associated EBV proteins expressed in the tumor cells provide potent antigenic targets for immunotherapeutic approaches *(1)*. Donor T cells have been used effectively to treat EBV–LPD *(6,7)*, but are associated with GvHD and when used to treat active disease, they can cause significant pathology due to inflammation. We show here that donor-derived, EBV-specific cytotoxic T lymphocytes (CTLs) can safely be given prophylactically and will prevent EBV–LPD without causing GvHD.

EBV is also found in the malignant cells of 50% of patients with Hodgkin's disease (HD) *(8–10)*, offering another potential target for CTL therapy. Although the long-term survival for the majority of patients with HD is good, the outlook for the 10% who fail initial therapy is poor. Furthermore, primary therapy for HD is aggressive and the incidence of second cancers and other long-term treatment-related toxic effects is high. Targeted immunotherapy may provide a less toxic alternative primary treatment, which should improve the quality of life of survivors and could offer salvage therapy to those who fail conventional therapy. Although EBV gene expression in the malignant cells of HD is more limited than in immunoblastic lymphoma cells, the three latency-associated proteins that are expressed (EBNA1, LMP1, and LMP2a) should provide antigenic targets for immunotherapeutic approaches *(2,11)*. We are currently analyzing the potential of EBV-specific CTL as therapy for relapsed HD.

Methods

Generation of EBV-Transformed B-Cell Lines

LCLs were prepared from bone marrow donors or patients with HD by infection of peripheral blood mononuclear cells (PBMCs) with concentrated viral supernatant from the B95-8 cell line *(12)*. Cells were cultured in flat-bottomed wells in the presence of 1 μg/mL of cyclosporin A until clumps of virus-transformed cells began to expand. After establishment, LCLs were maintained in acyclovir to prevent the production of infectious virus in the cultures *(12)*.

CTL Generation

EBV-specific CTL lines were generated by co-culture of 2×10^6 PBMCs with 5×10^4 irradiated autologous

**Affiliations of authors:* C. M. Rooney, M. A. Roskrow, C. A. Smith (Department of Virology and Molecular Biology), M. K. Brenner, H. E. Heslop (Department of Hematology and Oncology), St. Jude Children's Research Hospital, Memphis, TN.

Correspondence to: Cliona M. Rooney, Ph.D., Department of Pediatrics-Hematology/Oncology, Texas Children's Hospital, Baylor College of Medicine, 6621 Fannin St., Houston, TX 77030-2399. E-mail: cmrooney@msmail.his.tch.tmc.edu

See "Note" following "References."

© Oxford University Press

EBV-transformed B cells in 2-mL wells. CTLs were expanded by weekly restimulation with the irradiated autologous LCL and twice weekly stimulation with 20 U/mL of recombinant interleukin 2 after day 14. CTLs were gene marked with a retrovirus vector containing the bacterial neomycin resistance (neo) gene as described previously *(12)*.

Safety Testing of CTL Lines

When sufficient CTL numbers were achieved, the line was safety tested and frozen. Release criteria mandated that the line should possess the donor HLA type; should contain less than 1% B cells; should be free of contaminating bacteria, fungi, and mycoplasma; should have low endotoxin; and should have minimal cytotoxic activity against recipient-derived phytohemagglutinin-induced lymphoblasts (PHA blasts), a measure of potential GvHD activity. Supernatants from gene-marked lines were saved to test for replication-competent retrovirus.

CTL Assay

Target cells, which included autologous HLA class I-mismatched LCLs, a lymphokine-activated killer-sensitive T-cell lymphoma, HSB-2, and PHA blasts derived from the recipients prior to bone marrow transplant (BMT), were labeled with 100 mCi ^{51}Cr in a small volume of medium for 1 hour at 37 °C. They were then washed extensively and incubated with CTLs at the required effector : target ratio for 5 hours. Half the supernatant was then removed from each well and counted. The percent specific lysis was calculated using the standard formula,

$$\text{percent specific lysis} = \frac{\text{experimental release} - \text{medium release}}{\text{maximum release} - \text{medium release}}.$$

PCR for neo and EBV

DNA was prepared from PBMC using QIAmp Blood PCR kits (Qiagen, Chatsworth, CA). For quantitation of EBV, dilutions of DNA from 1 to 0.1 μg were subjected to a 25-cycle polymerase chain reaction (PCR) using 25 m*M* $MgCl_2$, 1 μ*M* of each primer, and 200 μ*M* deoxynucleotides. The primers, first described by Boyle et al. *(13)*, amplified a unique region of the EBNA2 gene that distinguishes between the two strains of EBV. The amplified products were electrophoresed on a 2.5% agarose gel, transferred to nylon MSI membranes, and detected by DNA hybridization using nonradiolabeled EBV probes according to the GENIUS (Boehringer Mannheim Corp, Indianapolis, IN) system *(14)*. Quantitation of neo sequences was carried out as described previously *(15)*. PCR products were electrophoresed, Southern blotted and probed with a radiolabeled neospecific probe. PCR products were compared with known standards and the sensitivity of detection was 0.01% neomarked cells.

Precursor CTL Assay

Patient PBMCs were plated in flat-bottomed 96-well plates at six to ten doubling dilutions from 5×10^5 cells per well with 10^4 cells from the irradiated autologous LCL and 10^5 irradiated autologous PBMC as feeders. Replicates ranged from 6 to 12 depending on the PBMC numbers available. Cells were restimulated weekly with the autologous irradiated LCL and fed twice weekly with 20 U per mL of interleukin 2 from day 14. Cytotoxicity against the autologous and an HLA-mismatched LCL was tested after 6 weeks of culture.

Dendritic Cell Preparation

Dendritic cells (DC) were enriched from peripheral blood by adhering 10^6 PBMC to 2 mL culture wells in 1 mL of medium for 2 hours. Nonadherent cells were removed gently and 1 mL of medium containing 800 U per mL granulocyte–macrophage colony-stimulating factor (GM–CSF) and 500 U per mL interleukin 4 was added to the adherent cells *(16)*. These cells were cultured for 3–5 days with the addition of extra GM–CSF on day 3. Between 20% and 50% of these cells had DC morphology and phenotype (CD3-, CD14-, CD19-, DR+, and CD1a+). The crude population was used for transduction.

Dendritic Cell Transduction

Cells (10^6) that were enriched for DC were transduced with a retrovirus vector containing the LMP2a gene, using a flow-through transduction method. Retrovirus supernatant (20–40 mL) was concentrated on an Anocell 25 membrane by applying a pressure of 40–60 mm Hg. DC in 1 mL were then gently adhered to the virus on the membrane by applying a pressure of 20 mm Hg for 5–10 minutes. The membrane was then placed in a 6-mL well and cultured in 5 mL of complete medium containing GM–CSF for 16 hours before analysis for retrovirus integration or for irradiation and use as stimulator cells.

Results and Discussion

Safety

Forty-five recipients of T-cell-depleted marrow have received donor-derived, EBV-specific CTL, generated by stimulation of donor PBMC with the autologous irradiated, EBV-transformed B cell line. Twenty-six EBV-specific CTL lines were gene marked prior to infusion to determine the *in vivo* fate of the infused cell line. All patients were treated on protocols approved locally by our Institutional Review Board and federally by the Recombinant DNA Advisory Committee of the National Institutes of Health and the Food and Drug Administration *(17)*.

CTL as prophylaxis. Forty-two patients received EBV-specific CTL as prophylaxis and in these patients there was no short- or long-term toxicity. Eleven patients died, eight from relapse of the original leukemia, two from bacterial sepsis unrelated to CTL infusion, and one from pneumonitis. One patient developed a recurrence of GvHD, with no accumulation of marked T cells in biopsy tissue, and three developed skin rashes attributed to drug sensitivity.

CTL as treatment. Three patients who did not receive prophylactic CTL (two were ineligible and one refused prophylaxis) received CTL as treatment for fulminant disease *(5,18)*. Disease regressed completely in two patients, one without incident, but one with severe morbidity as a result of the inflammatory response to his bulky and extensive disease. The third patient died of overwhelming disease, 25 days after CTL infusion.

Persistence

Twenty-six patients received CTL that had been gene marked with efficiencies ranging from 0.1% to 10%. Marked cells could be tracked by PCR analysis of peripheral blood DNA prepared from patients for up to 18 weeks after CTL infusion (Fig. 1) *(19)*. Semiquantitative analysis of marker DNA indicated an expansion of CTL *in vivo* of up to three logs in some patients. If

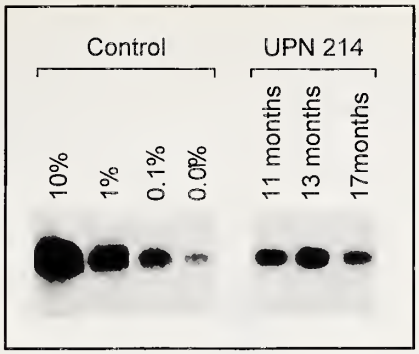


Fig. 1. Marked cells detected for up to 17 months in patient-derived EBV-specific CTL lines. Southern blot of neospecific PCR of DNA prepared from EBV-specific CTL lines initiated from patient UPN 214 at the indicated times after CTL infusion. Signals are compared with control cells in which from 0.01% to 10% of the cells contain one copy of the marker gene.

EBV-specific CTLs were selectively expanded from patients, marked T cells could be detected for up to 3 years postinfusion *(19)*. The generation of an immune response to the marker gene and resultant destruction of marked cells did not appear to be a problem in our patient group, since marked cells persisted long term in all but two assessable patients.

Immune Reconstitution

Evidence for immune reconstitution came from comparisons of the precursor frequency of EBV-specific T cells before CTL infusion and 1 and 4 weeks after. In 15 patients, there was a 35-fold median increase in precursor frequency (range, 1.5 to >500-fold) *(19)*. An example of the increase in precursor frequency is demonstrated in Fig. 2.

Antiviral Effects

We had previously shown that in stem cell transplant recipients, more than 2000 EBV genomes per 10^6 peripheral blood mononuclear cells were highly predictive of EBV–LPD *(14)*. Seven of 63 individuals who had high EBV DNA levels and did not receive CTL prophylaxis developed EBV–LPD. By contrast, while six patients in the prophylaxis group also developed high EBV DNA, in all of these individuals, EBV DNA levels dropped by two to four logs within 3 weeks of infusion, and none subsequently developed lymphoma *(19)*.

Antitumor Effects

Three lines of evidence indicated that infused EBV-specific CTLs had antitumor effects. First, none of 42 patients who received prophylactic CTL developed EBV–LPD compared with seven of 63 who did not receive CTL (either because they were transplanted before the study opened or because they were ineligible or refused). Second, none of the patients who developed high EBV DNA and received CTL developed EBV–LPD. The third line of evidence came from patients who responded to infusion of EBV-specific CTL as treatment for bulky disease. In one of these patients, we obtained biopsy specimens and found evidence for accumulation or expansion of infused CTL in tumor biopsy tissue. In this case, *in situ* PCR analysis demonstrated marked cells in tumor tissue biopsied 10 days after CTL infusion. Comparative analysis of the marker gene in peripheral blood and tumor tissue demonstrated selective accumulation or amplification of marked cells in the tumor tissue. About 1% of cells in the tumor tissue were marked by comparison with less than 0.01% in peripheral blood. Although this patient recovered completely and has been disease free for more than 1 year, he had severe morbidity as a result of the CTL-mediated inflammatory response to his diffusely invasive disease.

Prophylaxis is Better Than Cure

Although we successfully treated two of three patients with extensive lymphoma, a third patient who received CTL as treatment of advanced pulmonary disease died of respiratory failure with overwhelming tumor burden. Together with the morbidity experienced by the second patient, our results show that, while prophylactic use of EBV-specific CTL is safe and effective, treatment of this rapidly progressive and invasive disease by infusion of EBV-specific CTL can cause damage and may be ineffective in advanced cases.

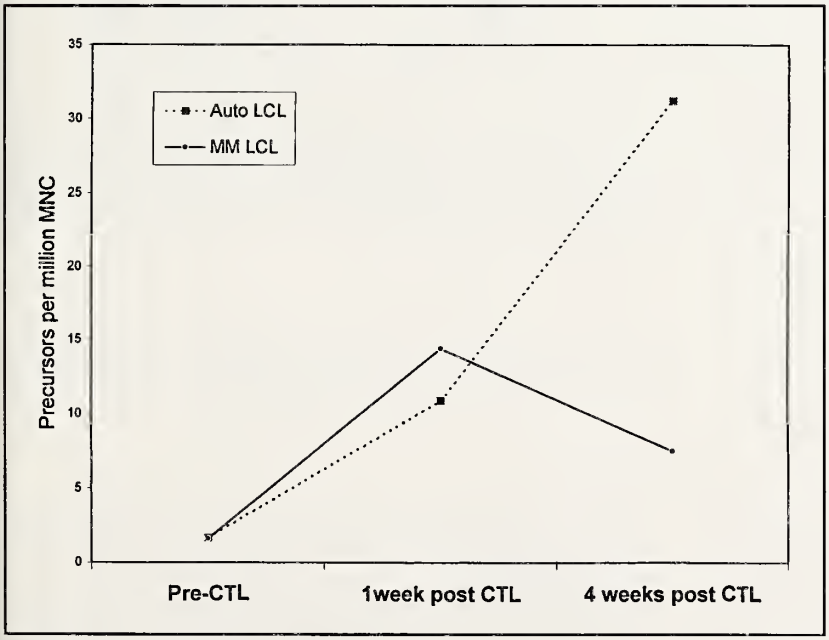


Fig. 2. The frequency of EBV-specific CTL precursors increases after CTL infusion. Here shown is an example of one patient in whom CTL precursors that kill the autologous LCL increase over the 4 weeks following infusion of one dose of 2×10^7 CTL/m^2. Dotted line shows the number of precursors that kill the autologous LCL and solid line shows killing of an HLA class I-mismatched LCL.

Treatment of EBV-Positive HD

To expand on this successful prevention of EBV–lymphoma in allogeneic stem cell transplant recipients, we decided to apply CTL to a second EBV-associated tumor, HD. Two protocols were designed and approved, one for the treatment of patients who have relapsed after therapy and have active disease and the second as adjuvant therapy for patients who receive autologous BMT as treatment for relapsed disease. EBV-specific CTL lines were activated and expanded *in vitro*, using the patient-derived EBV-transformed B cell line as antigen-presenting cells. CTLs were derived successfully from 9 of 13 patients, although lines grew more slowly from patients with HD than from normal donors and usually required additional mitogenic stimulation *(20)*.

A drawback to the approach of using LCL as the

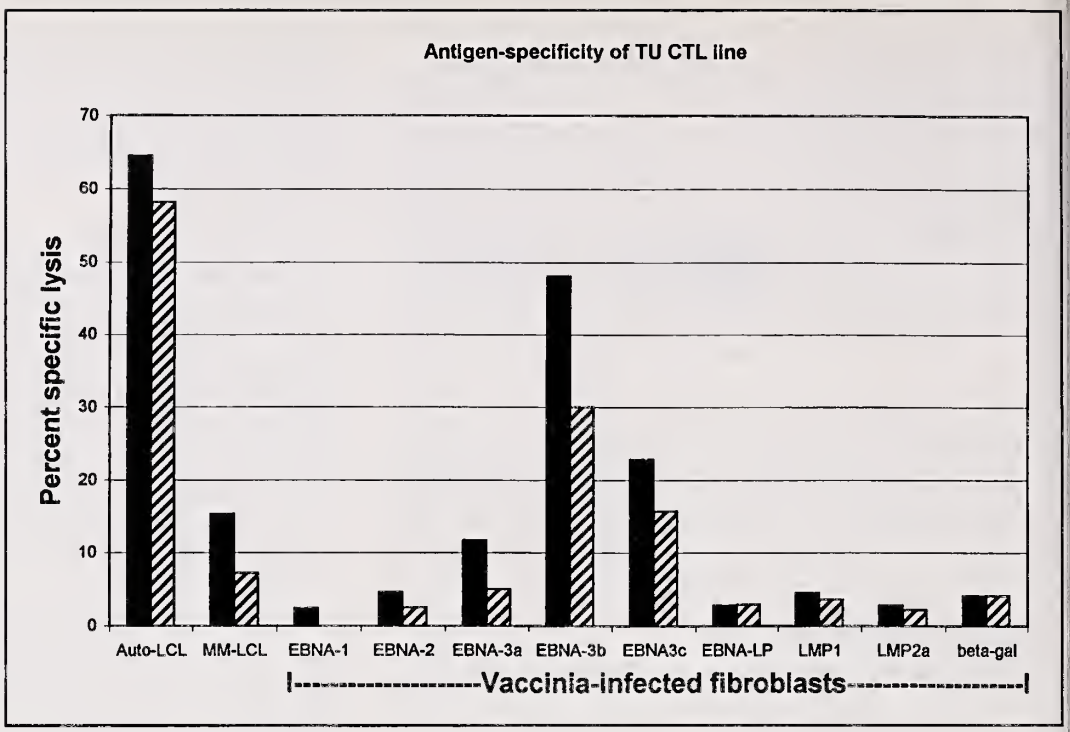


Fig. 3. EBV-specific CTLs generated by co-culture with the autologous EBV-transformed B cell line have limited antigen specificity. The specificity of the TU CTL line was tested against the autologous TU LCL and an HLA-mismatched LCL (MM LCL). The antigen specificity was tested using autologous TU fibroblasts that had been infected with vaccinia recombinants expressing the eight known EBV latency-associated proteins that are expressed in LCL. Only fibroblasts expressing EBNAs 3b and 3c are recognized. There is no significant killing of fibroblasts expressing EBNA1, LMP1, or LMP2a.

antigen-presenting cell for the generation of CTL for HD is that these CTL may not contain clones with specificity for the three EBV proteins expressed in Hodgkin's tumor cells. These three, EBNA-1, LMP1, and LMP2, are poorly immunogenic by comparison with EBNAs 3a, 3b, and 3c, which are immunodominant on the majority of HLA backgrounds *(2,11)*. Fig. 3 shows that CTLs generated by co-culture with LCL contain clones with specificity for a limited range of EBV proteins. If the CTL lines generated from patients with HD contained no clones with specificity for EBNA1, LMP1, or LMP2, then they would have no effect on disease.

To generate CTL with specificity for the proteins expressed in HD tumor cells from patients in whom these proteins may be poorly immunogenic, we are using DCs to present proteins individually. DCs are the most potent antigen-presenting cells known and have been shown both to induce primary immune responses *in vitro* and to overcome nonresponsiveness to tumor antigens *(21)*. Since EBNA1 is not processed appropriately for class I recognition *(22)* and LMP1 is heterogeneous between EBV isolates *(23)*, we have chosen to express LMP2a in DC for stimulation of autologous CTLs, using retrovirus vector containing the LMP2a complementary DNA. As shown in Fig. 4, LMP2a-transduced DCs are able to induce autologous CTLs that kill not only autologous fibroblasts expressing LMP2a from a vaccinia virus vector but also the autologous LCL that is a biologic target for EBV. Future studies will assess the efficacy of LMP2a-specific CTL in the treatment of EBV-positive HD.

In summary, we provide clinical evidence that infusion of EBV-specific CTL provides safe and effective prophylaxis for EBV-associated lymphoproliferative disease in stem cell transplant recipients. CTLs persist for up to 3 years *in vivo*, they rapidly restore cellular immune responses to EBV, and they reduce high EBV load to nonthreatening levels. As treatment for advanced and rapidly progressive disease, CTL may be less

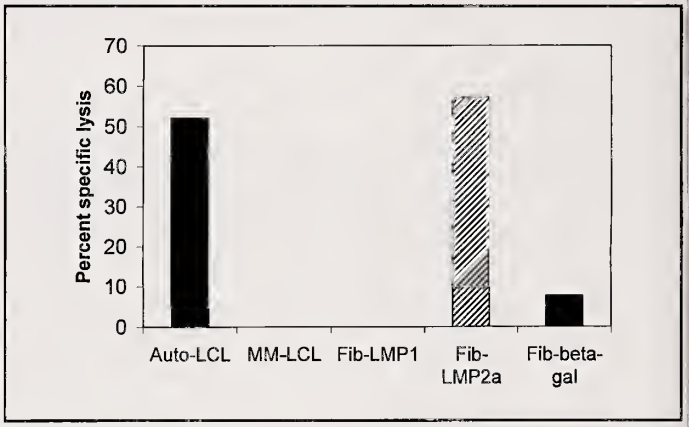


Fig. 4. CTLs induced by stimulation with autologous dendritic cells transduced with a retrovirus-LMP2a construct kill fibroblasts infected with a vaccinia LMP2a recombinant. Autologous fibroblasts are also killed strongly. HLA-mismatched LCL and fibroblasts infected vaccinia recombinants expressing LMP1 or β-galactosidase are not killed.

effective and may cause pathology. However, our studies provide good rationale for applying cellular immunotherapy to other cancers with identifiable target antigens.

References

(1) Rickinson AB, Kieff E. Epstein-Barr virus. In: Fields BN, Knipe DM, Howley PM, editors. Fields virology. Philadelphia: Lippincott-Raven, 1996, pp 2397–446.

(2) Rickinson AB, Lee SP, Steven NM. Cytotoxic T lymphocyte responses to Epstein-Barr virus. Curr Opin Immunol 1996;8:492–7.

(3) Straus SE, Cohen JI, Tosato G, Meier J. Epstein-Barr virus infections: biology, pathogenesis and management. Ann Intern Med 1992;118:45–58.

(4) O'Reilly RJ, Small TN, Papadopoulos E, Lucas K, Lacerda J, Koulova L. Biology and adoptive cell therapy of Epstein-Barr virus-associated lymphoproliferative disorders in recipients of marrow allografts. Immunolog Rev 1997;157:195–216.

(5) Heslop HE, Rooney CM. Adoptive immunotherapy of EBV lymphoproliferative diseases. Immunolog Rev 1997;157:217–22.
(6) Papadopoulos EB, Ladanyi M, Emanuel D, MacKinnon S, Boulad F, Carabasi MH, et al. Infusions of donor leukocytes to treat Epstein-Barr virus-associated lymphoproliferative disorders after allogeneic bone marrow transplantation. N Engl J Med 1994;330:1185–91.
(7) Heslop HE, Brenner MK, Rooney CM. Donor T cells to treat EBV-associated lymphoma. N Engl J Med 1994;331:679–80.
(8) Herbst H, Steinbrecher E, Niedobitek G, Young L, Brooks L, Muller-Lantzsch N, et al. Distribution and phenotype of Epstein-Barr virus-harboring cells in Hodgkin's disease. Blood 1992;80:484–91.
(9) Pallesen G, Hamilton-Dutoit SJ, Rowe M, Young LS. Expression of Epstein-Barr virus latent gene products in tumour cells of Hodgkin's disease. Lancet 1991;337:320–2.
(10) Gulley ML, Eagan PA, Quintanilla-Martinez L, Picardo AL, Smir BN, Childs C, et al. Epstein-Barr virus DNA is abundant and monoclonal in the Reed–Sternberg cells of Hodgkin's disease: association with mixed cellularity subtype and hispanic American ethnicity. Blood 1994;83:1595–602.
(11) Khanna R, Burrows SR, Moss DJ. Immune regulation in Epstein-Barr virus-associated diseases. Microbiolog Rev 1995;59:387–405.
(12) Smith CA, Ng CY, Heslop HE, Holladay MS, Richardson S, Turner EV, et al. Production of genetically modified EBV-specific cytotoxic T cells for adoptive transfer to patients at high risk of EBV-associated lymphoproliferative disease. J Hematother 1995;4:73–9.
(13) Boyle MJ, Sewell WA, Sculley TB, Apolloni A, Turner JJ, Swanson CE, et al. Subtypes of Epstein-Barr virus in human immunodeficiency virus-associated non-Hodgkin lymphoma. Blood 1991;78:3004–11.
(14) Rooney CM, Loftin SK, Holladay MS, Brenner MK, Krance RA, Heslop HE. Early identification of Epstein-Barr virus-associated post-transplant lymphoproliferative disease. Br J Haematol 1995;89:98–103.
(15) Brenner MK, Rill DR, Holladay MS, Heslop HE, Moen RC, Buschle M, et al. Gene marking to determine whether autologous marrow infusion restores long-term haemopoiesis in cancer patients. Lancet 1993;342:1134–7.
(16) Romani N, Gruner S, Brang D, Kampgen E, Lenz A, Trockenbacher B, et al. Proliferating dendritic cell progenitors in human blood. J Exp Med 1994;180:83–93.
(17) Heslop HE, Brenner MK, Rooney CM, Krance RA, Roberts WM, Rochester R, et al. Administration of neomycin-resistance-gene-marked EBV-specific cytotoxic T lymphocytes to recipients of mismatched-related or phenotypically similar unrelated donor marrow grafts. Hum Gene Ther 1994;5:381–97.
(18) Rooney CM, Smith CA, Ng C, Loftin SK, Li C, Krance RA, et al. Use of gene-modified virus-specific T lymphocytes to control Epstein-Barr virus-related lymphoproliferation. Lancet 1995;345:9–13.
(19) Heslop HE, Ng CY, Li C, Smith CA, Loftin SK, Krance RA, et al. Long-term restoration of immunity against Epstein-Barr virus infection by adoptive transfer of gene-modified virus-specific T lymphocytes. Nat Med 1996;2:551–5.
(20) Roskrow MA, Suzuki N, Heslop HE, Hudson M, Brenner MK, Rooney CM. Genetically modified EBV-specific cytotoxic T cells for adoptive transfer to patients with EBV-positive Hodgkin disease. Blood 1996;88:673a.
(21) Steinman RM, Witmer-Pack M, Inaba K. Dendritic cells: antigen presentation, accessory function and clinical relevance. Adv Exp Med Biol 1993;329:1–9.
(22) Levitskaya J, Coram M, Levitsky V, Imreh S, Steigerwald-Mullen PM, Klein G, et al. Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. Nature 1995;375:685–8.
(23) Khecht H, Bachmann E, Brousset P, Sandvej K, Nadal D, Bachmann F, et al. Deletions within the LMP1 oncogene of Epstein-Barr virus are clustered in Hodgkin's disease and identical to those observed in nasopharyngeal carcinoma. Blood 1993;82:2937–42.

Note

Supported in part by Public Health Service grants CA21765, CA81364, and CA71426 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services; by the Assisi Foundation of Memphis; and by the American Lebanese Syrian Associated Charities.

Genetic Basis of Acquired Immunodeficiency Syndrome-Related Lymphomagenesis

*Gianluca Gaidano, Antonino Carbone, Riccardo Dalla-Favera**

The molecular pathogenesis of systemic acquired immunodeficiency syndrome (AIDS)-related non-Hodgkin's lymphomas (AIDS–NHL) is a complex process involving both host factors and the accumulation of genetic lesions within the tumor clone. On the basis of the pattern of molecular lesions involved in these tumors, several distinct pathogenetic pathways can be presently identified in AIDS-related lymphomagenesis. These pathways selectively associate with the different clinicopathologic variants of AIDS–NHL. AIDS-related Burkitt's lymphoma is characterized by activation of c-MYC in all cases, disruption of p53 in 60% of the cases, and infection by Epstein-Barr virus (EBV) in 30% of the cases. AIDS-related diffuse large-cell lymphoma harbor frequent EBV infection (80%) and, in 20% of the cases, BCL-6 rearrangements. Finally, the pathogenesis of AIDS-related body cavity-based lymphoma involves infection by human herpesvirus-8 in all cases and frequently also the co-infection by EBV. [Monogr Natl Cancer Inst 1998;23:95–100]

Non-Hodgkin's lymphoma (NHL) is the second most frequent cancer associated with acquired immunodeficiency syndrome (AIDS) after Kaposi's sarcoma, and in some risk groups, namely, the hemophiliacs, NHL represent the most common AIDS-related neoplasm *(1,2)*. The incidence of AIDS-related NHLs (AIDS–NHL) has continued to rise since the outburst of the AIDS epidemic and, in 1985, the Centers for Disease Control (CDC) has recognized NHL as an AIDS-defining condition *(1–3)*. Because AIDS–NHL are frequently a late complication of AIDS, it is thought that their incidence will continue to increase steadily with the prolongation of life expectancy of human immunodeficiency virus (HIV)-infected individuals *(4)*.

All AIDS–NHL share a number of similarities *(2,5)*. They all derive from B cells, are characterized by extreme clinical aggressiveness, and display a predilection for extranodal sites, including unusual locations otherwise rarely implicated by B-cell NHL of the immunocompetent host. Despite these apparent similarities, however, it is now well established that AIDS–NHL display a marked clinicomorphologic heterogeneity. The detailed pathologic classification of AIDS–NHL has been a matter of controversy and has been recently updated by the World Health Organization *(6)*. First of all, AIDS–NHL may present as a systemic disease (systemic AIDS–NHL) or may originate and locate selectively in the central nervous system (primary central nervous system lymphomas, PCNSL). Depending on histology, AIDS–NHL are generally distinguished into two major categories, which include Burkitt's lymphoma (BL) and diffuse large cell lymphoma (DLCL). AIDS-related BL (AIDS–BL) and AIDS-related DLCL (AIDS–DLCL) account for approximately one third and two thirds of systemic AIDS–NHL, respectively *(1)*. Conversely, AIDS–PCNSL are consistently represented by the DLCL histology. An additional, although rare, pathologic category of AIDS–NHL is constituted by body cavity-based lymphoma (AIDS–BCBL), a peculiar type of lymphoma growing in liquid phase in the serous cavities of the body *(7–9)*.

The pathogenesis of AIDS–NHL is a complex process implicating different phases *(2)*. Initially, factors contributed by the immunocompromised host, and mainly represented by disrupted immunosurveillance and chronic antigen stimulation, lead to polyoligo-clonal B-cell hyperstimulation and hyperplasia, thus configurating the clinicopathologic picture known as persistent generalized lymphadenopathy (PGL) *(2,10)*. The role of disrupted immunosurveillance is exemplified by the close association between low levels of CD4-positive T cells in the host's peripheral blood and increased risk of AIDS–NHL development *(11)*. This association is particularly marked in the case of AIDS–DLCL and AIDS–PCNSL, whereas AIDS–BL may also develop in the context of a relatively preserved immune function. The contribution of chronic antigen stimulation and the selection to AIDS-related lymphomagenesis is documented by the high rate of somatic mutations accumulating in the hypervariable regions of immunoglobulin (Ig) genes expressed by AIDS–NHL cells as well as by the preferential usage of Ig-variable genes belonging to Ig gene families, such as V_H4, which are frequently implicated in the generation of B-cell autoreactive clones *(12)*.

Within this context of B-cell hyperstimulation and hyperplasia, the emergence and progressive expansion of the AIDS–NHL clone are driven by the clonal accumulation of genetic lesions within the neoplastic population. The search for these genetic lesions has been the focus of intense efforts during the last 5 years of research in this field. Here we report on the identification of the molecular alterations most frequently associated with the major categories of AIDS–NHL.

**Affiliations of authors:* G. Gaidano, Division of Internal Medicine, Department of Medical Sciences, University of Torino at Novara, Italy; A. Carbone, Division of Pathology, I.N.R.C.C.S., Centro di Riferimento Oncologico, Aviano, Italy; R. Dalla-Favera, Division of Oncology, Department of Pathology, College of Physicians & Surgeons, Columbia University, New York, NY

Correspondence to: Riccardo Dalla-Favera, M.D., Division of Oncology, Department of Pathology, College of Physicians & Surgeons, Columbia University, 630 W. 168th St., New York, NY 10032. E-mail: rd10@columbia.edu.

See "Note" following "References."

© Oxford University Press

Identification of Genetic Alterations in AIDS–NHL

The very first molecular studies performed on AIDS–NHL suggested that the genetic lesions associated with this group of neoplasms are markedly heterogeneous. These same studies demonstrated that, as in other human cancers, genetic lesions of AIDS–NHL may cause different consequences, including the activation of cellular proto-oncogenes, the disruption of tumor suppressor genes, or, alternatively, the insertion of foreign genes in the neoplastic cells, as exemplified by the case of tumor infection by oncogenic viruses.

To define with precision the patterns of genetic lesions associated with the various categories of AIDS–NHL, we have investigated extensively the molecular features of large panels of systemic AIDS–NHL derived from different geographic backgrounds, including the United States and Northern Italy. We have chosen to focus our attention on the genetic lesions most frequently involved in aggressive B-cell NHL of the immunocompetent host, including the alterations of c-MYC and BLC-6 among proto-oncogenes, the inactivation of p53 and 6q among tumor suppressor loci, and the infection by oncogenic herpesviruses.

Viral Infection

Several studies *(2,5)* have focused on the involvement of EBV in AIDS–NHL. These studies have demonstrated that the virus infects the tumor clone of a large proportion of systemic AIDS–NHL, including 30% AIDS–BL and 70%–80% AIDS–DLCL *(13–15)*. EBV infection is also detectable in 100% AIDS–PCNSL and approximately 90% AIDS–BCBL *(7–9,16)*. In most AIDS–NHL tested, infection by EBV has been demonstrated to be monoclonal, consistent with a model of infection preceding, and putatively contributing to, clonal expansion of the neoplastic population *(13,14)*. Among EBV-positive AIDS–NHL, the expression pattern of the EBV-encoded transforming antigens EBNA-2 and LMP-1 varies substantially *(17)*. Cases of AIDS–BL are consistently negative for expression of both antigens, whereas systemic AIDS–DLCL express LMP-1 in a substantial fraction of cases, approximately 50% *(17,18)*. Some rare cases of systemic AIDS–DLCL also express EBNA-2.

In addition to EBV, another herpesvirus, namely, human herpesvirus type-8 (HHV-8), is involved in AIDS–NHL pathogenesis, although at a substantially lower frequency *(7–9)*. Infection of the tumor clone by HHV-8 is restricted to cases of AIDS–BCBL, whereas it is consistently absent in all other AIDS–NHL types *(7–9,19)* (Fig. 1, A–C). Other viruses, including HIV and HTLV-I, do not appear to be directly implicated in AIDS–NHL pathogenesis, since they do not infect the tumor clone *(2)*.

c-MYC

Since the initial phases of the AIDS–NHL epidemic, cytogenetic studies had revealed substantial similarities between AIDS–BL and BL of the immunocompetent host, based on the presence of chromosomal translocations affecting band 8q24, the site of the c-MYC proto-oncogene, and an Ig locus *(20)*. These initial suggestions were later confirmed by molecular analysis of c-MYC in AIDS–NHL samples *(14)*. These genetic studies aimed at defining the frequency and distribution of c-MYC alterations throughout the spectrum of AIDS–NHL and at characterizing the precise mechanism of activation of the proto-oncogene in these tumors. Analysis of large panels of cases have revealed that activation of c-MYC associated selectively with all cases of AIDS–BL, whereas it is absent among AIDS–DLCL and AIDS–BCBL *(7–9,14)*. Although rare cases of AIDS–DLCL have been reported to harbor an activated c-MYC, these samples presumably represent cases of AIDS–BL that have been misdiagnosed because of the well-known histologic pleomorphism characteristic of AIDS–NHL *(21)*.

Regarding the precise mechanism of c-MYC activation in AIDS–BL, it generally reflects that typical of sporadic BL, as opposed to endemic BL *(14)*. Thus, in the case of t(8;14)(q24;q32), the most frequent translocation type affecting c-MYC, the translocation breakpoints on chromosome 8 fall within sequences internal or immediately 5′ to the c-MYC gene, whereas the chromosome 14 breakpoints map to the Ig switch region *(22)*. As a consequence of the translocation, the expression of the c-MYC gene undergoes transcriptional deregulation *(22)*. In addition to being truncated, the c-MYC gene in AIDS–BL is also frequently mutated within its exon 2 *(23)*. These mutations lead to amino acid substitutions of the c-MYC protein and alter the physiologic interactions between c-MYC and its partner proteins, namely p107 *(24)*. Whereas in normal conditions p107 is able to suppress the transactivation properties of c-MYC, mutated c-MYC alleles are no longer responsive to p107 and thus escape one of the main physiologic mechanisms regulating c-MYC activity *(24)*.

BCL-6

The BCL-6 gene has been originally cloned by virtue of its involvement in chromosomal translocations affecting chromosomal band 3q27 in the DLCL of the immunocompetent host *(25)*. Among NHL of the immunocompetent host, BCL-6 is rearranged in approximately 40% of B-cell DLCL *(26)*. Among AIDS–NHL, rearrangements of BCL-6 cluster selectively with a fraction of AIDS–DLCL, whereas they are absent among AIDS–BL and AIDS–BCBL *(27,28)*. The frequency of BCL-6 rearrangements in AIDS–DLCL is significantly lower than that detected among DLCL of the immunocompetent host, confirming the notion that the pathogenesis of these two groups of neoplasms is different (*see also* the p53 section below).

Rearrangements involving BCL-6 are promiscuous, in that several distinct chromosomal sites may serve as chromosomal partners juxtaposing to 3q27 *(25)*. In contrast to c-MYC translocations, the chromosomal partners of 3q37 include, but are not restricted to, the Ig loci. The rearrangement breakpoints cluster within a 4-kb region spanning the BCL-6 promoter sequences and the first noncoding exons and result in the fusion of BCL-6 coding sequences (exons 2–10) to heterologous promoters derived from the partner chromosome. This mechanism of BCL-6 alteration is termed promoter substitution and is thought to cause deregulated expression of the BCL-6 protein *(29)*.

In addition to chromosomal rearrangements, the BCL-6 gene in AIDS–NHL may be altered by an alternative type of genetic alteration. Recent studies of NHL of the immunocompetent host have shown that in approximately 70% DLCL and 50% follicular lymphomas the BCL-6 gene is altered by multiple mutations

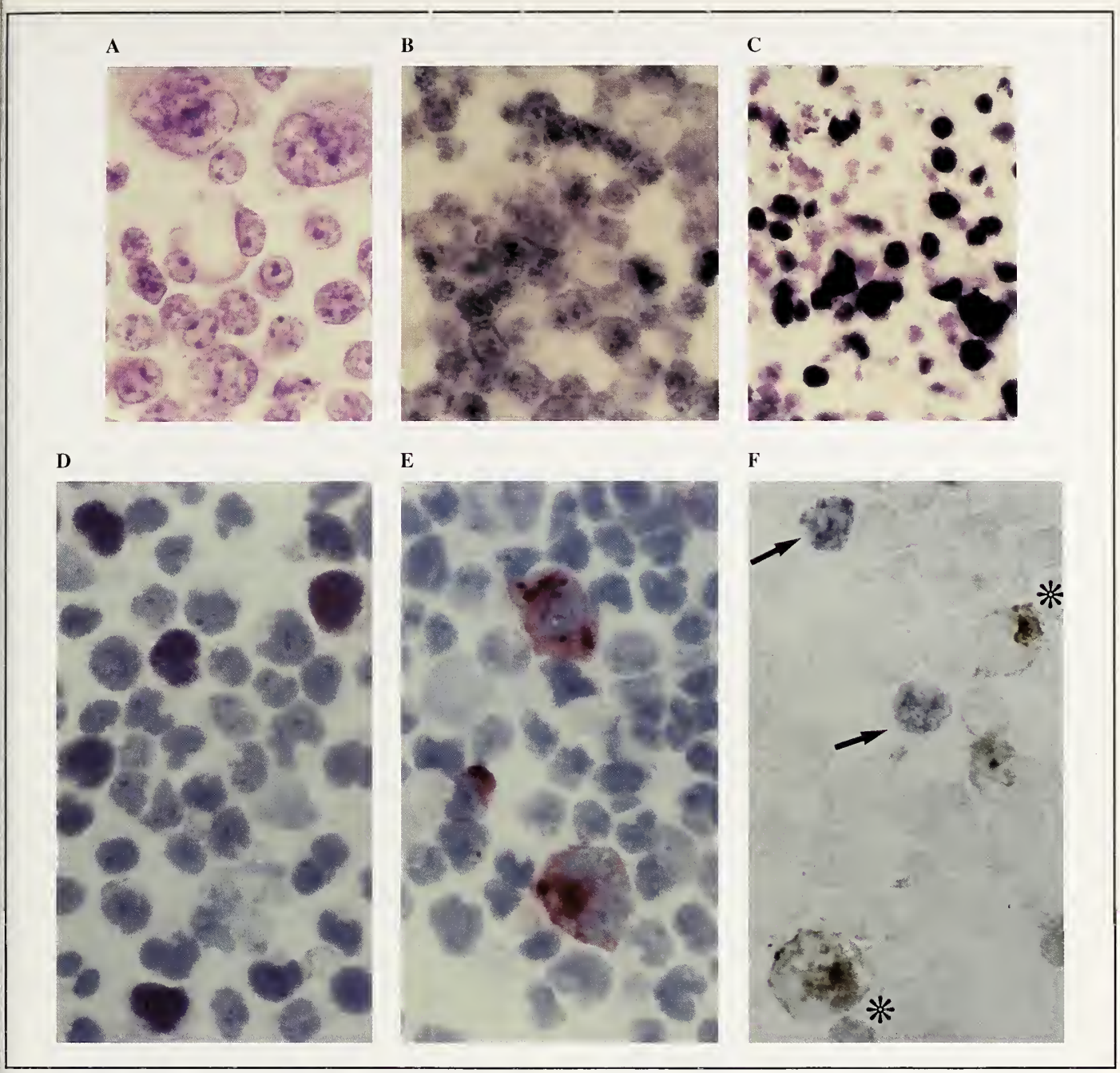


Fig. 1. A) AIDS-related HHV-8$^+$ body-cavity-based lymphoma. In a cell block from pleural effusion, the tumor cells are characterized by nuclei that are irregularly shaped and variably chromatic with prominent, often multiple nucleoli. Most tumor cells have moderately abundant cytoplasm with plasmacytoid appearance. Hematoxylin–eosin stain, original magnification ×630. **B)** AIDS-related HHV-8$^+$ body-cavity-based lymphoma-derived cell line (HBL-6). HHV-8 *in situ* hybridization signal is present as dense grains over the nuclei of most tumor cells. Cytospin preparation; digoxigenin-labeled KS330 probe; *in situ* hybridization, nuclear fast red counterstain, original magnification ×630. **C)** AIDS-related body-cavity-based lymphoma. In a cell block from pleural effusion, EBER *in situ* hybridization signal is present as dense grains over the nuclei of most tumor cells. *In situ* hybridization, nuclear fast red counterstain, original magnification ×250. **D)** HBL-2/EBV obtained by EBV *in vitro* infection of HBL-2, an EBV-negative AIDS–BL cell line. Nuclear positivity for BCL-6 is restricted to a small number of tumor cells. Cytospin preparation, APAAP method, original magnification ×400. **E)** HBL-2/EBV obtained by EBV *in vitro* infection of HBL-2, an EBV-negative AIDS–BL cell line. LMP-I cytoplasmic positivity is detected on two tumor cells of large size. Cytospin preparation, APAAP method, original magnification ×400. **F)** HBL-2/EBV obtained by EBV *in vitro* infection of HBL-2, an EBV-negative AIDS–BL cell line. Double labeling for nuclear BCL-6 (blue) and cytoplasmic LMP-1 (brown). Against a background of unstained cells, individual tumor cells express either BCL-6 (arrows) or LMP-1 (asterisks). No coexpression of both proteins by the same tumor cell is detectable. Cytospin preparation, original magnification ×400.

clustering in its 5′ noncoding region *(30)*. These mutations frequently occur in the absence of any recognizable chromosomal abnormality affecting band 3q27 or molecular rearrangement of the BCL-6 locus. The genomic sequences most frequently involved by these mutations are adjacent to the BCL-6 promoter region and overlap with the major cluster of chromosomal breakpoints, suggesting that mutations and rearrangements may be selected for their ability to alter the same region. The combined frequency of mutations and rearrangements approaches 100% of DLCL cases arising in the immunocompetent host, suggesting that structural alterations of the 5′ noncoding region of the BCL-6 gene are necessary for the development of these tumors

(30). To verify the involvement of BCL-6 mutations among AIDS–NHL, we have performed an extensive analysis of this genetic lesion by a combination of PCR–SSCP and DNA direct sequencing in a panel representative of the major pathologic categories of these lymphomas. These data show that mutations of the 5′ noncoding regions of BCL-6 are detectable in approximately 60% of systemic AIDS–NHL *(28)*. According to histology, mutations were found to cluster with approximately 70% AIDS–BL and AIDS–DLCL, whereas they were restricted to 20% AIDS–BCBL *(28)*. Within the AIDS–DLCL group, BCL-6 5′ mutations occurred both in the presence and in the absence of BCL-6 rearrangements, thus mimicking the pattern reported for DLCL of the immunocompetent host. With respect to other genetic lesions frequently encountered in AIDS–NHL, the presence of BCL-6 5′ mutations in a given AIDS–NHL sample was independent of the concomitant presence of c-MYC rearrangements, p53 mutations, and EBV infection *(28)*. The extreme frequency of BCL-6 mutations renders these alterations the most frequent genetic lesion of systemic AIDS–NHL. Since BCL-6 mutations are not only frequent, but are also specific for a given tumor sample, they may be exploited as a molecular marker for monitoring the course of the disease.

In addition to exploring the genetic features of BCL-6 in AIDS–NHL, we have recently investigated the expression status of the BCL-6 protein in these tumors. The BCL-6 protein belongs to the family of transcription factors containing zinc-finger motifs *(25)*. Functional studies have indicated that BCL-6 can function as a transcriptional repressor that can bind a specific DNA sequence and repress transcription from linked promoters *(31)*. Thus, the physiologic function of BCL-6 may be to repress the expression of genes carrying its specific DNA binding motif. Several observations indicate that BCL-6 is involved in the development and function of the germinal center (GC). First, the BCL-$6^{-/-}$ phenotype associates with lack of GC formation, as demonstrated by animal models in which the expression of the physiologic BCL-6 protein has been constitutively abrogated in the germline *(32)*. In addition, since the BCL-6 gene is expressed in GC B cells, but not in their differentiated cell progenies (plasma cells and memory B cells), it is conceivable that BCL-6 may be involved in the induction and sustainment of GC-associated functions and that BCL-6 down-regulation may be necessary for B cells to progress toward further differentiation into memory B cells or plasma cells *(33)*. In NHL carrying a rearranged BCL-6, the down-regulation of the BCL-6 protein may be prevented by the juxtaposition of the gene to heterologous promoters.

Among AIDS–NHL, expression of the BCL-6 protein was found in 100% AIDS–BL and in approximately 50% AIDS–DLCL *(18)*. Intriguingly, among AIDS–DLCL, expression of the BCL-6 protein was mutually exclusive with expression of the EBV-encoded antigen LMP-1 *(18)*. The precise molecular mechanism for this phenomenon is presently unclear, although it is interesting to note that BCL-6^{+}/LMP-1^{-} AIDS–DLCL generally display a large noncleaved cell morphology, whereas cases of BCL-6^{-}/LMP-1^{+} AIDS–DLCL are classifiable as immunoblastic plasmacytoid lymphomas *(18)*. We are currently investigating whether this phenotypic heterogeneity of AIDS–DLCL reflects a different clinical behavior of these two AIDS–DLCL variants.

The concept that LMP-1 and BCL-6 are mutually exclusive antigens was also based on experimental data derived from the comparison of the AIDS–BL cell line HBL-2 and its EBV-infected counterpart HBL-2/EBV. The EBV-negative HBL-2 cell line expresses high levels of the BCL-6 protein, whereas the cell line HBL-2/EBV, which expresses LMP-1 in a fraction of cells, exhibits a marked reduction in the number of BCL-6-positive tumor cells. Notably, no co-expression of BCL-6 and LMP-1 can be detected in a single tumor cell of HBL-2/EBV (Fig. I, D–F).

The results of genetic and protein studies of BCL-6 may also help us to understand the histogenesis of AIDS–NHL. On one hand, in fact, mutations of BCL-6 5′ regulatory regions are generally regarded as a marker of transition of a given B cell through the GC *(30)*. Since a large fraction of these tumors does carry BCL-6 mutations *(28)*, it is conceivable that they derive from B-cell subsets that have already transited through (and potentially reside in) the GC. Furthermore, expression of the BCL-6 protein is also considered as a marker of a B-cell differentiation state corresponding to B cells residing in the GC *(33)*. In this respect, all AIDS–BL and a fraction of AIDS–DLCL may thus be derived from GC-proliferating B cells, i.e., the centroblasts *(18)*. According to this same model, the fraction of AIDS–DLCL that do not express BCL-6, but conversely express LMP-1 when EBV infected, may derive from more differentiated B cells that have already exited from the GC *(18)*.

p53

Inactivation of the p53 tumor suppressor gene represents one of the most frequent genetic alterations in human cancers *(22)*. Among B-cell neoplasms of the immunocompetent host, mutations of p53 are virtually restricted to the case of BL and DLCL transformed from a previous follicular phase *(22)*. Inactivation of p53 in human neoplasia most commonly occurs through mutation of one allele and deletion of the other allele, although biallelic deletions, or, alternatively, bi-allelic mutations, may also occur in some cases. In the case of AIDS–NHL, our analysis of more than 50 cases by PCR–SSCP and DNA direct sequencing has demonstrated that p53 inactivation clusters selectively with AIDS–BL, whereas it is consistently absent among cases of AIDS–DLCL and AIDS–BCBL *(8,14,28)*. The pattern of p53 inactivation in AIDS–NHL reflects that of human neoplasia in general. The frequency of p53 mutations among AIDS–BL (60%) far exceeds that of non-AIDS-related BL, including both sporadic and endemic BL (30%) *(14)*. The reason for this excess of p53 mutations in AIDS–BL is presently unknown, although one hypothesis is that it may be related to the host's immunodeficit. Finally, the absence of p53 mutations among AIDS–DLCL, combined to the negativity for BCL-2 rearrangements of these tumors, is consistent with the *de novo* origin of AIDS–DLCL. In fact, among DLCL of the immunocompetent host, cases arising *de novo* are devoid of BCL-2 and p53 lesions, which, conversely, denote cases arising from a preceding follicular lymphoma *(22)*.

6q Deletions

Deletions of the long arm of chromosome 6 (6q) represent a frequent genetic alteration in B-cell NHL of the immunocom-

petent host *(22)*. Although the relevant tumor suppressor gene is presently unknown, it is thought to map to a small region of chromosomal band 6q27. Among AIDS–NHL, deletions of 6q27 cluster with a fraction of AIDS–DLCL (20%), whereas deletions are consistently absent in other AIDS–NHL types *(34)*. Deletions of 6q27 in AIDS–DLCL may occur in combination with other genetic alterations typical of this lymphoma category, including BCL-6 rearrangements and EBV infection.

Conclusions

The data summarized in this report demonstrate that AIDS–NHL is a strikingly heterogeneous disease. At present, four major molecular pathways can be identified. Each of these molecular pathways associates with peculiar clinical features and is restricted to a given AIDS–NHL histologic type. The first pathway associates with AIDS–BL and is characterized by relatively mild immunodeficiency of the host and multiple genetic lesions of the tumor, including activation of c-MYC, disruption of p53, and, although less frequently, infection by EBV. Typically, EBV-infected AIDS–BL fail to express the viral transforming antigens LMP-1 and EBNA-2.

Two distinct pathways associate with AIDS–DLCL, a type of AIDS–NHL generally characterized by a marked disruption of immune function. Whereas the majority of AIDS–DLCL carry EBV infection, only a fraction of cases express the viral antigen LMP-1. Expression of LMP-1 and BCL-6 segregate the two pathways associated with AIDS–DLCL. On the one hand, in fact, LMP-1-positive AIDS–DLCL fail to express the BCL-6 protein and display features consistent with immunoblastic-plasmacytoid differentiation, suggesting a derivation from post-GC cells. On the other hand, LMP-1-negative AIDS–DLCL express BCL-6 and display a large noncleaved cell morphology, suggesting an origin from the GC.

Finally, the fourth pathway associates with AIDS–BCBL. This rare lymphoma type consistently harbors infection by HHV-8 and, frequently, also by EBV. All other genetic lesions commonly detected among AIDS–NHL are consistently negative in AIDS–BCBL.

References

(1) Beral V, Petterman T, Berkelman R, Jaffe H. AIDS-associated non-Hodgkin's lymphoma. Lancet 1991;337:805–9.

(2) Gaidano G, Dalla-Favera R. Molecular pathogeneis of AIDS-related lymphomas. Adv Cancer Res 1997;67:113–53.

(3) Center for Disease Control. Revision of the case definition of acquired immunodeficiency syndrome for national reporting—United States. Ann Intern Med 1985;103:402–8.

(4) Gail MH, Pluda JM, Rabkin CS, Biggar RJ, Goedert JJ, Horm JW, et al. Projection of the incidence of non-Hodgkin's lymphoma related to acquired immunodeficiency syndrome. J Natl Cancer Inst 1991;83:695–700.

(5) Gaidano G, Carbone A. AIDS-related lymphomas: from pathogenesis to pathology. Br J Haematol 1995;90:235–43.

(6) IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Human Immunodeficiency Viruses and Human T-cell Lymphotropic Viruses. France, 1–18 June 1996. IARC Monogr Eval Carcinog Risks Hum 1996;67:1–424.

(7) Cesarman E, Chang Y, Moore PS, Said JW, Knowles DM. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. N Engl J Med 1995;332:1186–91.

(8) Carbone A, Globhini A, Vaccher E, Zagonel V, Pastore C, Dalla Palma P, et al. Kaposi's sarcoma-associated herpesvirus DNA sequences in AIDS-related and AIDS-unrelated lymphomatous effusions. Br J Haematol 1996;94:533–43.

(9) Carbone A, Gaidano G. HHV-8-positive body-cavity-based lymphoma: a novel lymphoma entity. Br J Haematol 1997;97:515–22.

(10) Pelicci PG, Knowles DM, Arlin ZA, Wieczorek R, Luciw P, Dina D, et al. Multiple monoclonal B cell expansions and c-myc rearrangements in acquired immune deficiency syndrome-related lymphoproliferative disorders: implications for lymphomagenesis. J Exp Med 1986;164:2049–76.

(11) Pluda JM, Venzon D, Tosato G, Lietzau J, Wyvill K, Nelson DL, et al. Parameters affecting the development of non-Hodgkin's lymphoma in patients with severe human immunodeficiency virus infection receiving antiretroviral therapy. J Clin Oncol 1993;11:1099–107.

(12) Riboldi P, Gaidano G, Schettino EW, Steger TG, Knowles DM, Dalla-Favera R, et al. Two AIDS-associated Burkitt's lymphomas produce specific anti-i IgM cold agglutinins utilizing somatically mutated V_H4-21 segments. Blood 1994;83:952–61.

(13) Neri A, Barriga F, Inghirami G, Knowles DM, Neequaye J, Magrath IT, Dalla-Favera R. Epstein-Barr virus infection precedes clonal expansion in Burkitt's and acquired immunodeficiency associated-lymphoma. Blood 1991;77:1092–5.

(14) Ballerini P, Gaidano G, Gong JZ, Tassi V, Saglio G, Knowles DM, Dalla-Favera R. Multiple genetic lesions in acquired immunodeficiency syndrome-related non-Hodgkin's lymphoma. Blood 1993;81:166–76.

(15) Carbone A, Gloghini A, Zanette I, Canal B, Volpe R. Demonstration of Epstein-Barr virus genomes by in situ hybridization in acquired immune deficiency syndrome-related high grade and anaplastic large cell $CD30^+$ lymphomas. Am J Clin Pathol 1993;99:289–97.

(16) MacMahon EME, Glass JD, Hayward SD, Mann RB, Becker PS, Charache P, McArthur JC, Ambinder RF. Epstein-Barr virus in AIDS-related primary central nervous system lymphoma. Lancet 1991;338:969-73.

(17) Carbone A, Tirelli U, Gloghini A, Volpe R, Boiocchi M. Human immunodeficiency-virus-associated systemic lymphomas may be subdivided into two main groups according to Epstein-Barr viral latent gene expression. J Clin Oncol 1993;11:1674–81.

(18) Carbone A, Gaidano G, Gloghini A, Pastore C, Saglio G, Tirelli U, et al. BCL-6 protein expression in AIDS-related non-Hodgkin's lymphomas. Am J. Pathol 1997;150:155–65.

(19) Gaidano G, Pastore C, Gloghini A, Cusini M, Nomdedeu J, Volpe G, et al. Distribution of human herpesvirus-8 sequences throughout the spectrum of AIDS-related neoplasia. AIDS 1996;10:941-9.

(20) Chaganti RS, Jhanwar SC, Koziner B, Arlin Z, Mertelsmann R, Clarkson B. Specific translocations characterize Burkitt's like lymphoma of homosexual men with the acquired immunodeficiency syndrome. Blood 1983;61:1265–8.

(21) Delecluse HJ, Raphael M, Magaud JP, Felman P, the French Study Group of Pathology for Human Immunodeficiency Virus-associated tumors, Alsamad IA, et al. Variable morphology of human immunodeficiency virus-associated lymphomas with c-MYC rearrangements. Blood 1993;82:552–63.

(22) Gaidano G, Dalla-Favera R. Molecular biology of lymphomas. In: DeVita VT, Hellman S, Rosenberg SA, editors. Cancer: principles and practice of oncology. Philadelphia: Lippincott, 1997:2131–45.

(23) Bhatia K, Spangler G, Gaidano G, Hamdy N, Dalla-Favera R, Magrath I. Mutations in the coding region of c-myc occur frequently in acquired immunodeficiency syndrome-associated lymphomas. Blood 1994;84:883–8.

(24) Gu W, Bhatia K, Magrath IT, Dang CV, Dalla-Favera R. Binding and suppression of the Myc transcriptional activation domain by p107. Science 1994;264:251–4.

(25) Ye BH, Lista F, LoCoco F, Knowles DM, Chaganti RS, Dalla-Favera R. Alterations of blc-6, a novel zinc-finger gene, in diffuse large cell lymphoma. Science 1993;262:747–50.

(26) LoCoco F, Ye BH, Lista F, Corradini P, Offit K, Knowles DM, et al. Rearrangements of the bcl-6 gene in diffuse large cell non-Hodgkin's lymphoma. Blood 1994;83:1757–9.

(27) Gaidano G, LoCoco F, Ye BH, Shibata D, Levine A, Knowles DM, et al. Rearrangements of the bcl-6 gene in acquired immunodeficiency syndrome-associated non-Hodgkin's lymphoma: association with diffuse large-cell subtype. Blood 1994;84:397–402.

(28) Gaidano G, Carbone A, Pastore C, Capello D, Migliazza A, Gloghini A, et al. Frequent mutation of the 5′ noncoding region of the BCL-6 gene in acquired immunodeficiency syndrome-related non-Hodgkin's lymphomas. Blood 1997;89:3755–62.

(29) Ye BH, Chaganti S, Chang CC, Niu H, Corradini P, Chaganti RS, et al. Chromosomal translocations cause deregulated BCL6 expression by promoter substitution in B cell lymphoma. EMBO J 1995;14:6209–17.

(30) Migliazza A, Martinotti S, Chen W, Fusco C, Ye BY, Knowles DM, et al. Frequent somatic hypermutation of the 5′ noncoding region of the bcl-6 gene in B-cell lymphoma. Proc Natl Acad Sci USA 1995;92:12520–4.

(31) Chang CC, Ye BH, Chaganti RS, Dalla-Favera R. BLC-6, a POZ/zinc-finger protein, is a sequence-specific transcriptional repressor. Proc Natl Acad Sci USA 1996;93:6947–52.
(32) Ye BH, Cattoretti G, Shen Q, Zhang J, Hawe N, de Waard R, et al. The BCL-6 proto-oncogene controls germinal-centre formation and Th2-type inflammation. Nat Genet 1997;16:161–70.
(33) Cattoretti G, Chang CC, Cechova K, Zhang J, Ye BH, Falini B, et al. BCL-6 protein is expressed in germinal-center B cells. Blood 1995;86:45–53.
(34) Pastore C, Carbone A, Gloghini A, Volpe G, Saglio G, Gaidano G. Association of 6q deletions with AIDS-related diffuse large cell lymphoma. Leukemia 1996;10:1051–3.

Note

Supported by Public Health Service grant CA37295 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services; by Ministero della Sanità, Istituto Superiore di Sanità, AIDS project 1997, Rome, Italy; and by Fondazione "Piera, Pietro e Giovanni Ferrero," Alba, Italy.

Clinical Management of Human Immunodeficiency Virus-Associated Non-Hodgkin's Lymphoma

*Lawrence D. Kaplan**

The human immunodeficiency virus (HIV)-associated non-Hodgkin's lymphomas are best divided into three categories: 1) primary central nervous system (CNS) lymphomas, 2) systemic lymphomas, and 3) primary effusion lymphomas. As described elsewhere in this monograph, these lymphomas differ in many of their molecular characteristics and presumed mechanisms of pathogenesis. These factors may be relevant in terms of development of future therapeutic approaches that take advantage of some of these unique molecular characteristics. However, of more immediate relevance is the fact that these lymphomas occur in different patient populations and are associated with different clinical outcomes. Systemic lymphomas are observed across a broad range of levels of immune function, with a median $CD4^+$ cell count of approximately $100/mm^3$ *(1,2)*. The median survival among patients with systemic lymphoma is approximately 5–8 months, and 10%–20% survive disease free for longer than 2 years *(2–4)*. In contrast, primary CNS lymphoma occurs in the most severely immunocompromised group, 75% of whom have CD4 counts less than $50/mm^3$ *(1)*. CNS lymphoma in the setting of this profound immunosuppression is associated with a median survival of only 3 months. The diagnosis and management of primary CNS lymphoma therefore present unique clinical management problems, and it is here that I will begin the discussion of clinical management.

Primary CNS Lymphomas

Diagnosis

The difficulty in diagnosing primary CNS lymphoma is related to the fact that lymphoma in the brain is difficult to distinguish from intercerebral toxoplasmosis, a common opportunistic infection in patients with HIV disease. Although the radiographic appearance on computed tomography or magnetic resonance imaging scan may provide clues as to the diagnosis, these studies are rarely diagnostic in acquired immunodeficiency syndrome (AIDS) patients with CNS lesions. Since magnetic resonance scanning in individuals with toxoplasmosis reveals a solitary lesion in only 21% of cases *(5)*, the finding of a solitary lesion is more suggestive of a diagnosis of lymphoma. However, magnetie resonance scanning in individuals with CNS lymphoma demonstrates multiple lesions in 50% of cases *(5)*, and this finding is less likely to be helpful. The diagnosis of lymphoma may also be suggested by a periventricular location or a lesion that crosses the midline. However, these "classic" radiographic findings occur infrequently in HIV-associated CNS lymphoma.

Recently, thallium-201 single-photon emission computed tomography (SPECT) scanning has been used to distinguish between these two entities. In one study *(6)*, all 24 individuals with toxoplasmosis had negative scans, whereas all 12 with CNS lymphoma had positive scans. In another study *(7)*, nine patients with CNS lymphoma were all positive on thallium-201 SPECT scanning. Of concern, however, was the fact that three of 10 patients with toxoplasmosis also had positive scans. Although these series of patients are small and data from larger numbers of patients are needed, these observations do suggest a potential utility for this nuclear medicine study in the diagnosis of CNS lymphoma.

Of even greater interest are observations made with the use of polymerase chain reaction (PCR) for detecting Epstein-Barr virus (EBV) DNA sequences in cerebrospinal fluid. De Luca et al. *(8)* showed that seven of eight individuals with documented CNS lymphoma had positive PCR for EBV in cerebrospinal fluid. All 11 patients with brain lesions and no lymphoma were negative, and 21 individuals with AIDS but no CNS lesions were also negative. In a second series *(9)*, all 17 individuals with CNS lymphoma were positive for EBV by PCR and 67 of 68 individuals with HIV and no lymphoma were negative. Data from a much larger series of patients have been reported elsewhere in this monograph and confirm these results. These data strongly suggest that this technique is a useful means of noninvasive diagnosis of primary CNS lymphoma in patients with HIV disease. This may be particularly powerful when used in conjunction with thallium-201 SPECT scanning and the results of *Toxoplasma* serologic studies.

At the present time, the standard of care for diagnosis of CNS lymphoma is brain biopsy. Since treatment outcome strongly depends on early diagnosis and institution of therapy, it is strongly recommended that individuals who are *Toxoplasma* seronegative, and therefore unlikely to have a diagnosis of toxoplasmosis, be referred for early brain biopsy if cerebrospinal fluid cytology is negative.

Once the diagnosis of CNS lymphoma has been established, all patients should undergo a slit lamp examination before initiation of therapy to rule out the presence of ocular lymphoma.

**Affiliations of author:* Department of Medicine, San Francisco General Hospital, and Department of Medicine, University of California.

Correspondence to: Lawrence D. Kaplan, M.D., University of California, San Francisco, San Francisco General Hospital/AIDS Program, Oncology Division, Ward 84, 995 Potrero Ave., San Francisco, CA 941110.

© Oxford University Press

Treatment

In most series reported in the literature, patients have been treated with a course of whole-brain radiotherapy. In three series reporting the occurrence of clinical improvement with therapy *(10–12)*, 62%–79% had improvement in their neurologic symptoms. Overall response rates of 52%–70% have been reported, with complete response rates of 40%–50% *(1,10–14)*. Median survival has been 2–4.8 months, with most patients dying as a result of nonlymphoma complications from their HIV disease *(1,10–14)*. However, the proportion of treated patients dying as a result of lymphoma progression ranged from 13% to 55%, suggesting that improvement in treatment of the lymphoma itself is clearly needed *(1,10–14)*.

Two small pilot studies *(15,16)* have investigated the use of combined modality therapy with chemotherapy and radiotherapy being administered. In one series *(15)*, 10 individuals treated in this fashion had a complete response rate of 88%, and none of these patients died as a result of their CNS lymphoma. However, the reported median survival of 3.5 months was no better than that reported in studies using whole-brain radiotherapy alone. It is clear then that improvement in the management of CNS lymphoma will require not only improvement in management of the lymphoma, but also better treatment for the underlying HIV disease. The current use of multiagent antiretroviral therapies, which have been shown to be highly effective, may prolong survival from the underlying immunodeficiency disease long enough so that better management of the lymphoma may soon become a more critical factor.

Systemic Lymphomas

Although primary CNS involvement is associated with the worst prognosis of all presentations of AIDS-associated lymphoma, a number of factors have been associated with clinical outcome in systemic disease. Early retrospective studies *(1,3)* demonstrated factors associated with the underlying immunodeficiency disorder such as CD4 count, presence or absence of a prior AIDS-defining diagnosis, and performance status to be the most important predictors of outcome. Presence or absence of extranodal disease, particularly bone marrow involvement, was the only tumor-related factor associated with prognosis.

Data from more recent studies, such as the prospective data on 192 patients enrolled in the ACTG 142 study *(17)*, a study of low-dose versus standard-dose chemotherapy, more closely resemble those factors identified in the International Prognostic Index *(18)* for non-immunodeficiency-associated aggressive non-Hodgkin's lymphomas. In this study *(17)*, factors associated with poor prognosis included age older than 35 years, CD4 count of less than 100/mm^3, history of intravenous drug use, and, for the first time, tumor bulk as measured by stage of disease. Advanced stage III or IV disease was associated with a poor outcome. Similarly, a recent retrospective review of 96 patients treated for aggressive HIV-associated lymphomas *(19)* indicated that bulky tumor, as measured by elevation of serum lactate dehydrogenase levels, was again associated with a poor clinical outcome.

Treatment

A number of therapeutic approaches have been utilized in the management of AIDS-associated lymphoma. Initially, dose-intensive regimens were used in an effort to improve outcome. These regimens included high-dose methotrexate and high-dose cytarabine *(20)* or regimens containing high-dose cyclophosphamide *(3)*. In general, these treatments were associated with a high risk of death due to opportunistic infection and survival times of 5–6 months, which were not better than those seen with more standard-dose therapy. Patients treated in these studies had median CD4 counts in the mid-100 range.

A large clinical trial of the aggressive LNH-84 regimen *(21)* targeted individuals with higher CD4 counts (median count, 227/mm^3) and was associated with a high complete remission rate, a median survival time of 9 months, and a 2-year disease-free survival of 42%. However, it cannot be demonstrated in this trial whether the improved clinical outcome is due to the aggressive chemotherapy regimen or to the fact that patients with higher CD4 counts are more likely to survive longer.

Because of the disappointing results initially observed with higher dose regimens, studies of dose-reduced chemotherapy regimens were subsequently conducted. The AIDS Clinical Trials Group *(22)* utilized an mBACOD regimen (methotrexate, bleomycin, doxorubicin, cyclophosphamide, vincristine, and dexamethasone) in which the dosages of cyclophosphamide and doxorubicin were reduced by approximately 50% in 35 patients who had aggressive AIDS-associated lymphomas. The complete response rate of 46% and the median survival of 5.6 months were not significantly different from those observed in previous trials of standard-dose chemotherapy, but they were achieved at the expense of significantly less hematologic toxicity. In this study *(22)*, the median CD4 count was 150/mm^3. Subsequently, an Italian study of a similar low-dose regimen in 37 patients with lymphoma and a median CD4 count of 25/mm^3 *(23)* achieved only a 14% complete remission rate and a median survival of only 3.5 months. Taken together, these data imply a more important effect of immune function on survival than choice of therapeutic regimen.

In view of these observations, the AIDS Clinical Trials Group designed a clinical trial to compare directly the importance of chemotherapy dose intensity for clinical outcome in AIDS-associated lymphoma. In this trial, patients were randomly assigned to receive either standard-dose mBACOD chemotherapy with granulocyte–macrophage colony-stimulating factor (GM-CSF) support or reduced-dose mBACOD with GM-CSF administered only as required for neutropenia *(2)*. The results (Table 1) demonstrated no significant difference in complete response rate, response duration, time to progression, and overall or disease-free survival. And the difference for response duration very nearly reached statistical significance in favor of the low-dose

Table 1. Standard- versus low-dose chemotherapy for acquired immunodeficiency syndrome-associated non-Hodgkin's lymphoma

Parameter	Dose: Standard	Dose: Low	P*
Complete response	42/81 (52%)	39/94 (41%)	NS
Time to progression	30 wk	39 wk	NS
Time to recurrence	106 wk	190 wk	.06
Median survival	31 wk	35 wk	NS
Toxicity ≥ grade 3	70% patients	51% patients	.008
Grade 4 neutropenia	39% cycles	24% cycles	.001

regimen. When subjects were divided into those with greater than or equal to or less than 100 CD4 cells/mm^3, those with better immune function did survive longer than those with poor immune function. However, within each of these two CD4 cohorts, no significant differences were observed with respect to treatment assignment, indicating that these results may apply to the majority of individuals with HIV-associated lymphoma and that the most appropriate choice of chemotherapy regimens for most individuals with AIDS-associated lymphoma should be a dose-reduced treatment regimen. However, an insufficient number of individuals with CD4 counts greater than or equal to 200/mm^3 were enrolled in this clinical trial to determine whether some of those patients may have benefited from the use of a standard-dose regimen. For this reason, it is recommended that some patients with relatively intact immune function (CD4$\geq$200/mm^3) be considered for standard-dose therapy.

Recently, the use of continuous infusion of cyclophosphamide, doxorubicin, and etoposide has been studied for the management of AIDS-associated lymphoma *(24)*. This treatment approach, piloted by Sparano et al., used a 96-hour continuous infusion of these three agents with granulocyte colony-stimulating factor. Patients assigned to group A received adjunctive therapy with didanosine in cycles 1, 2, 5, and 6; those assigned to group B received this antiretroviral agent in cycles 3, 4, 5, and 6. The complete response rate observed in this study of 56% was not significantly different from that observed in previous trials; however, the median survival time of 18.4 months reported in this study appeared to be significantly longer than that previously reported. It remains to be seen whether this longer survival time is due to improvement in antiviral therapy, more recent improvement in the overall management of HIV disease, or a regimen that truly produces more durable responses. A large phase II study of this chemotherapy regimen is currently under way by the Eastern Cooperative Oncology Group.

For those individuals who have refractory lymphoma, treatment outcome is particularly poor. Little data have been reported in the literature. Tirelli et al. *(25)* recently reported on the use of a combination of etoposide, mitoxantrone, and prednimustine in 21 patients with either primary resistance or relapse after having complete responses. In this study group, the complete response rate was 26%, and the median survival time was only 2 months; the majority of patients ultimately died of refractory lymphoma.

Data from San Francisco General Hospital (Kaplan LD: unpublished data), in which 14 patients were treated with escalating doses of infusional ifosfamide and etoposide, demonstrated a relatively similar outcome. The overall complete response rate in this series of patients was 43% with a relatively short median response duration of only 79 days.

Levine et al. *(26)* recently reported on the treatment of 35 patients with refractory AIDS-associated lymphoma with the single agent mitoguazone; although this agent appeared to be extremely well tolerated and non-myelosuppressive, the overall response rate of 23% suggests that mitoguazone may be better suited for use in combination regimens than as a single agent.

Meningeal Lymphomas

The reported incidence of lymphomatous meningitis at the time of diagnosis of systemic lymphoma ranges from 3% to 25%. Most of these cases are reported from relatively small retrospective series of patients. In the largest prospective series reported, the ACTG 142 study of 192 patients, the incidence of meningeal lymphoma at diagnosis was only 3%, suggesting that the frequency of this complication may be lower than initially suspected *(2)*. The frequency of meningeal relapse is a much more difficult number to determine because most reported series in the literature routinely treat their patients prophylactically with the use of intrathecal chemotherapy. However, in one small, early study, Gill et al. *(20)* treated 22 patients with a high-dose methotrexate, high-dose cytarabine treatment regimen. Despite the fact that these are agents that should have adequate CNS penetration, the frequency of meningeal relapse was very high (35%). However, it is noteworthy that seven of the eight relapses occurred in individuals who had bone marrow involvement at the time of their initial diagnosis of lymphoma. Similarly, in the series by Lowenthal et al. *(27)*, two of three individuals developing meningeal relapse had bone marrow involvement at diagnosis.

At the San Francisco General Hospital, we no longer routinely treat prophylactically all patients with systemic AIDS-associated lymphoma for meningeal recurrence. Rather, we target those individuals with risk factors for meningeal disease identified in the non-immunodeficiency-associated lymphoma population. These risk factors include bone marrow involvement, epidural disease, paranasal sinus involvement, and small, non-cleaved histologic pattern.

Antiviral Therapy

At this point in time, there is extensive experience in combining chemotherapy with nucleoside analogues. With the exception of concerns regarding overlapping myelosuppression when zidovudine (AZT) is used (and its use is therefore not recommended), there is little concern regarding the safety of these antiviral agents while chemotherapy is being administered. There are, however, little data concerning potential drug interactions between the protease inhibitors and agents such as cyclophosphamide and the anthracyclines. Clinical trials are currently investigating these potential drug interactions. Despite this lack of data, we believe that it is important to continue patients on their multiagent antiretroviral therapies while they are being administered chemotherapy. Physicians are urged to follow their patients closely for evidence of excessive toxicity when such combinations are used.

Future Therapeutic Approaches

There are undoubtedly a variety of mechanisms by which lymphoma arises in individuals with HIV disease. Lymphoma appears to arise out of a background of polyclonal B-cell hyperproliferation, which may arise through clonal integration of HIV in macrophages, resulting in overexpression of interleukin 6 or as a result of chronic antigenic stimulation from other sources (including viruses or other mitogens) *(28)*. Overexpression of interleukin 10 also appears to be involved in an autocrine fashion *(28,29)*. Other events that may involve viruses, such as EBV or human herpesvirus type 8 (HHV8), or genetic events, such as P53 mutations or c-myc gene rearrangements, may ultimately give rise to frank lymphoma.

At the same time, the cellular immune response to lymphoma

may be blocked by a variety of mechanisms, including high concentrations of interleukin 6 and interleukin 10, which may prevent CD8 cytotoxic lymphocytes from becoming activated *(28)*. The responsiveness of EBV-specific cytotoxic lymphocytes and natural killer (NK) cells appears to be blunted as well.

In vitro data as well as some preliminary *in vivo* data suggest that cytokines such as interleukin 2 *(30)* and interleukin 12 may be capable of improving CD8 and NK cell responses *(31)*, and clinical trials investigating the use of each of these agents as potential therapeutic modalities or as an adjunct to standard chemotherapy are currently under way. In addition, expansion of EBV-specific cytotoxic lymphocytes used in adoptive immunotherapy is another therapeutic modality that is currently being explored.

Unique molecular characteristics of AIDS-associated lymphomas can also be taken advantage of with the use of monoclonal antibodies and immunotoxins. Substantial antitumor responses to an anti-CD22-ricin-A-chain immunotoxin have been reported *(32)*, suggesting that this may be a potentially viable approach to therapy in some patients.

Anti-B4-blocked ricin has also been studied in individuals with refractory HIV-associated lymphoma with some observed responses, and clinical trials are currently in development for the use of anti-CD20 monoclonal antibodies, including the chimeric anti-CD20 antibody C2B8 which has been used successfully in the management of refractory low-grade lymphoma.

Primary Effusion Lymphomas

The primary effusion lymphomas constitute only approximately 5% of all HIV-associated lymphomas. These lymphomas present as body cavity effusions (pleural, pericardial, or ascites) that are generally not associated with a contiguous mass lesion and tend to remain localized in the body cavity of origin *(33)*. Morphologically, these lymphomas have a unique histologic pattern—one that is midway between that of large-cell immunoblastic and large-cell anaplastic lymphomas *(33)*. These lymphomas are unique in their characteristic association with HHV8 *(33,34)*. In addition, EBV genome was reported to be present in 100% of 15 tumors analyzed in one series *(33)* and in 50% of those evaluated in a second series of eight patients *(35)*. Both B- and T-cell-associated antigens are usually absent *(33)*.

Clinically, the majority of reported cases have occurred in homosexual or bisexual men *(33)*. The median CD4 lymphocyte count in one study (n = 19) was 84/mm^3; in a second study, it was 78/mm^3 *(35)*. Median survivals have been short (2–5 months), and no long-term disease-free survivors have been reported *(33,35)*. However, the number of such cases reported in the medical literature is small, and there may be individuals whose clinical outcome is significantly better than those outcomes reported.

We are beginning to learn a great deal about the biology and pathogenesis of the HIV-associated non-Hodgkin's lymphomas. It is hoped that this new knowledge will lead to the development of more pathogenesis-based approaches to therapy and to a greater ability to take advantage of some of the unique molecular features of these lymphomas so that therapies can be identified but will allow us to reduce the amount of cytotoxic therapy required for the management of this disease.

References

(1) Levine AM, Sullivan-Halley J, Pike MC, Rarick MU, Loureiro C, Bernstein-Singer M, et al. Human immunodeficiency virus-related lymphoma: prognostic factors predictive of survival. Cancer 1991;68:2466–72.

(2) Kaplan LD, Straus DJ, Testa MA, Von Roenn J, Dezube BJ, Cooley TP, et al. Randomized trial of standard-dose vs. low-dose mBACOD chemotherapy for HIV-associated non-Hodgkin's lymphoma. N Engl J Med 1997; 336:1641–8.

(3) Kaplan L, Abrams DI, Feigal E, McGrath M, Kahn J, Neville P, et al. AIDS-associated non-Hodgkin's lymphoma in San Francisco. JAMA 1989; 261:719–24.

(4) Kaplan LD, Northfelt D, Kahn J, Volberding P. Degree of immunodeficiency at diagnosis and treatment outcome in HIV-associated non-Hodgkin's lymphoma (HIV-NHL) [abstract]. Presented at 7th International Conference on AIDS, Amsterdam, July 19–24, 1992. Vol 2, p B105:# PoB3113.

(5) Ciricillo S, Rosenblum M. Use of CT and MR imaging to distinguish intracranial lesions and to define the need for biopsy in AIDS patients. J Neurosurg 1990;73:720–4.

(6) Ruiz A, Ganz WI, Post MJ, Camp A, Landy H, Maliin W, et al. Use of thallium-201 brain SPECT to differentiate cerebral lymphoma from toxoplasma encephalitis in AIDS patients. Am J Neuroradiol 1994;15:1885–94.

(7) Barker DE, Trepashko D, DeMarais P, Egofske P, Petry M, Langer B, et al. Utility of thallium brain SPECT in the exclusion of CNS lymphoma in AIDS [abstract]. Presented at 4th Conference on Retroviruses and Opportunistic Infections, Washington, DC, January 22–26, 1997, p 195:# 708.

(8) De Luca A, Antinori A, Cingolani A, Larocca LM, Linzalone A, Ammassari A, et al. Evaluation of cerebrospinal fluid EBV-DNA and IL-10 as markers for *in vivo* diagnosis of AIDS-related primary central nervous system lymphoma. Br J Haematol 1995;90:844–9.

(9) Cinque P, Brytting M, Vago L, Castagna A, Parravicini C, Zanchetta N, et al. Epstein-Barr virus DNA in cerebrospinal fluid from patients with AIDS-related primary lymphoma of the central nervous system. Lancet 1993;342: 398–401.

(10) Baumgartner J, Rachlin J, Beckstead J, Meeker TC, Levy RM, Wara WM, et al. Primary central nervous system lymphomas: natural history and response to radiation therapy in 55 patients with acquired immunodeficiency syndrome. J Neurosurg 1990;73:206–11.

(11) Nisce L, Kaufmann T, Metroka C. Radiation therapy in patients with AIDS-related central nervous system lymphomas. JAMA 1992;267: 1921–2.

(12) Lee J, Kaplan LD, Conant M, Northfelt DW. Natural history, treatment outcome and causes of death in AIDS-related primary central nervous system lymphoma (PCNSL). Proc ASCO 1994,18:54.

(13) Rubio R, Rubio M, Grauss F, Moreno S, Ribera JM, Pintado V, et al. Primary central nervous system lymphoma (PCNSL) in AIDS. A multicentric clinical study [abstract]. Presented at 11th International Conference on AIDS, Vancouver, July 7–12, 1996. Vol 2, p 291:# Th.B.4232.

(14) Ling SM, Roach M 3rd, Larson DA, Wara WM. Radiotherapy of primary central nervous system lymphoma in patients with and without human immunodeficiency virus. Ten years of treatment experience at the University of California, San Francisco. Cancer 1994;73:2570–82.

(15) Forsyth PA, Yahalom J, DeAngelis LM. Combined-modality therapy in the treatment of primary central nervous system lymphoma in AIDS. Neurology 1994;44:1473–9.

(16) Chamberlain MC. Long survival in patients with acquired immune deficiency syndrome-related primary central nervous system lymphoma. Cancer 1994;73:1728–30.

(17) Straus D, Huang J, Testa M, Levine A, Kaplan L. Prognostic factors in the treatment of HIV-associated non-Hodgkin's lymphoma: analysis of ACTG 142 (low-dose vs. standard-dose mBACOD + GM-CSF). Seattle (WA): American Society of Hematology, December 1–5, 1995, p 604a:# 2404.

(18) A predictive model for aggressive non-Hodgkin's lymphoma. The international non-Hodgkin's lymphoma prognostic factors project. N Engl J Med 1993;329:987–94.

(19) Vaccher E, Tirelli U, Spina M, Talamini R, Errante D, Simonelli C, et al. Age and serum lactate dehydrogenase level are independent prognostic factors in human immunodeficiency virus-related non-Hodgkin's lymphomas: a single-institute study of 96 patients. J Clin Oncol 1996;14:2217–23.

(20) Gill P, Levine AM, Krailo M, Rarick MU, Loureiro C, Deyton L, et al. AIDS-related malignant lymphoma: results of prospective treatment trials. J Clin Oncol 1987;5:1322–8.

(21) Gisselbrecht C, Oksenhendler E, Tirelli U, Lepage E, Gabarre J, Farcet J, et al. Human immunodeficiency virus-related lymphoma treatment with intensive combination chemotherapy. Am J Med 1993;95:188–96.

(22) Levine AM, Wernz JC, Kaplan L, Rodman N, Cohen P, Metroka C, et al. Low-dose chemotherapy with central nervous system prophylaxis and zidovudine maintenance in AIDS-related lymphoma. JAMA 1991;266: 84–8.

(23) Errante D, Tirelli U, Oksenhendler E, Vaccher E, Gastaldi R, Rizzardini G, et al. Prospective study with combined low-dose chemotherapy and zidovudine for 37 patients with poor prognosis HIV-related non-Hodgkin's lymphoma. Proc ASCO 1992;11:44.

(24) Sparano JA, Wiernik PH, Hu X, Sarta C, Schwartz EL, Soeiro R, et al. Pilot trial of infusional cyclophosphamide, doxorubicin, and etoposide plus didanosine and filgrastim in patients with human immunodeficiency virus-associated non-Hodgkin's lymphoma. J Clin Oncol 1996;14: 3026–35.

(25) Tirelli U, Errante D, Spina M, Gastaldi R, Nigra E, Nosari AM, et al. Second-line chemotherapy in human immunodeficiency virus-related non-Hodgkin's lymphoma: evidence of activity of a combination of etoposide, mitoxantrone, and prednimustine in relapsed patients. Cancer 1996;77: 2127–31.

(26) Levine AM, Tulpule A, Tessman D, Kaplan L, Giles F, Luskey BD, et al. Mitoguazone therapy in patients with refractory or relapsed AIDS-related lymphoma: results from a multicenter phase II trial. J Clin Oncol 1997;15: 1094–103.

(27) Lowenthal DA, Straus DJ, Campbell SW, Gold JW, Clarkson BD, Koziner B. AIDS-related lymphoid neoplasia: the Memorial Hospital experience. Cancer 1988;61:2325–37.

(28) Herndier BG, Kaplan LD, McGrath MS. Pathogenesis of AIDS lymphoma. AIDS 1994;8:1025–49.

(29) Emilie D, Touitou R, Raphael M, Peuchmaur M, Devergnee O, Rea D, et al. *In vivo* production of interleukin-10 by malignant cells in AIDS lymphomas. Eur J Immunol 1992;22:2937–42.

(30) Bernstein ZP, Porter MM, Gould M, Lipman B, Bluman EM, Stewart CC, et al. Prolonged administration of low-dose interleukin-2 in human immunodeficiency virus-associated malignancy results in selective expansion of innate immune effectors without significant clinical toxicity. Blood 1995; 86:3287–94.

(31) McFarland EJ, Harding PA, Schooley RT, Kuritzkes DR. Interleukin-12 (IL-12) enhances HIV-specific cytotoxic T-lymphocyte (CTL) activity [abstract]. Presented at 3rd Conference on Retroviruses and Opportunistic Infections, Washington, D.C., January 28–February 1, 1996. Vol 1, p 95:# 228.

(32) Kaplan LD, Robles R, Kahn JO, Schindler J, Ghetie V, Vitetta ES. Phase I trial of anti-CD22 ricin-A-chain immunotoxin (IgG-RFB4-dgA) for treatment of AIDS- lymphoma [abstract]. The National AIDS Malignancies Conference, April 28–30, 1997, Bethesda, MD.

(33) Nador RG, Cesarman E, Chadburn A, Dawson DB, Ansari MQ, Sald J, et al. Primary effusion lymphoma: a distinct clinicopathologic entity associated with the Kaposi's sarcoma-associated herpes virus. Blood 1996;88: 645–56.

(34) Nador RG, Cesarman E, Knowles DM, Said JW. Herpes-like DNA sequences in a body-cavity-based lymphoma in an HIV-negative patient. N Engl J Med 1995;33:943.

(35) Komanduri KV, Luce JA, McGrath MS, Herndier BG, Ng VL. The natural history and molecular heterogeneity of HIV-associated primary malignant lymphomatous effusions. JAIDS 1996;13:215–26.

Genetic Basis of Acquired Immunodeficiency Syndrome-Related Lymp
Gianluca Gaidano, Antonino Carbone, Riccardo Dalla-Favera

Clinical Management of Human Immunodeficiency Virus-Associated Non-Hodgkin's Lymphoma 101
Lawrence D. Kaplan

NIH LIBRARY
3 1496 00771 8623

MONOGRAPHS
JOURNAL OF THE NATIONAL CANCER INSTITUTE

Integrating Economic Analysis Into Cancer Clinical Trials: the National Cancer Institute–American Society of Clinical Oncology Economics Workbook

1998
Number 24

Contents

Introduction 1
Part I: Economic Analysis and Cancer Clinical Trials 1
Part II: Planning an Economic Analysis 5
Part III: Implementing an Economic Analysis 8
Part IV: Designing an Effective Data Collection Strategy 13
Part V: Statistical Issues 15
Part VI: Economic Evaluation in the Cooperative Group Setting 22
Conclusion 26
References 26
Notes 28

MONOGRAPHS

JOURNAL OF THE NATIONAL CANCER INSTITUTE

Number 24 *1998*
ISSN 0027-8874
ISBN 019-922378-5

LIBRARY
AUG 0 6 1998
National Institutes of Health

EDITORIAL BOARD

Barnett S. Kramer
Editor-in-Chief

J. Gordon McVie
European Editor

Eric J. Seifter
Book Reviews Editor

J. Paul Van Nevel
News Editor

Frederic J. Kaye
Douglas L. Weed
Reviews Editors

Martin L. Brown
Economics Editor

ASSOCIATE EDITORS

Susan G. Arbuck
Frank M. Balis
William J. Blot
Peter M. Blumberg
John D. Boice, Jr.
Louise A. Brinton
Bruce A. Chabner
Ross C. Donehower
Susan S. Ellenberg
Suzanne W. Fletcher
Michael A. Friedman
Patricia A. Ganz
Edward L. Giovannucci
John K. Gohagan
Frank J. Gonzalez
Michael M. Gottesman
Peter Greenwald
Donald E. Henson
Susan M. Hubbard
Colin R. Jefcoate
Frederic J. Kaye
Hynda K. Kleinman
Theodore S. Lawrence
Bernard Levin
W. Marston Linehan
Marc E. Lippman
Scott M. Lippman
Darrell T. Liu
Dan L. Longo
Reuben Lotan
Douglas R. Lowy
Susan G. Nayfield
David L. Nelson
Kenneth Olden
David G. Poplack
Ross L. Prentice
Alan S. Rabson
Edward A. Sausville
Robert H. Shoemaker
David Sidransky
Richard M. Simon
Michael B. Sporn
Maryalice Stetler-Stevenson
J. Paul Van Nevel
Douglas L. Weed
Noel S. Weiss

STATISTICAL EDITORS

Janet W. Andersen
Jacques Benichou
Donald A. Berry
Barry W. Brown
Bernard F. Cole
Susan S. Ellenberg
Scott S. Emerson
Eric Feuer
Edmund A. Gehan
Barry I. Graubard
Sylvan B. Green
Susan G. Groshen
Richard A. Olshen
Philip C. Prorok
Timothy R. Rebbeck
Philip S. Rosenberg
Harland N. Sather
Daniel J. Schaid
Richard M. Simon
Donald M. Stablein
Robert E. Tarone
Sholom Wacholder

EDITORIAL ADVISORY BOARD

Hans-Olov Adami
Harry B. Burke
Pelayo Correa
Ethan Dmitrovsky
Merrill J. Egorin
Soldano Ferrone
Isaiah J. Fidler
Jean L. Grem
Curtis C. Harris
Emily L. Harris
Kathy J. Helzlsouer
Maureen M. Henderson
Gloria H. Heppner
Allan Hildesheim
Waun Ki Hong
William J. Hoskins
Peter J. Houghton
David H. Johnson
V. Craig Jordan
Ilan R. Kirsch
Kenneth H. Kraemer
Alexandra M. Levine
Martha S. Linet
Henry T. Lynch
Frank L. Meyskens
Anthony B. Miller
Malcolm S. Mitchell
David L. Page
Kenneth J. Pienta
David Schottenfeld
Richard K. Severson
Jerry W. Shay
Sandra M. Swain
G. Marie Swanson
Mario Sznol
Raymond Taetle
Ian Tannock
Peter R. Twentyman

Richard D. Klausner
Director, National Cancer Institute
Susan Molloy Hubbard
Director, International Cancer Information Center
Julianne Chappell
Chief, Scientific Publications Branch; Executive Editor

EDITORIAL STAFF

Scientific Editors: Martin Brown, Ph.D.
Mary McCabe, R.N.
Kevin A. Schulman, M.D.

Monograph Coordinator: Elizabeth Horowitz

Manuscript Editors: Elaine Price Beck
Elizabeth Horowitz

MARKETING STAFF

Marketing Director: Jean Griffin Baum

Press Contact: Dan Eckstein

EDITORIAL POLICY: Manuscripts from key conferences dealing with cancer and closely related research fields, or a related group of papers on specific subjects of importance to cancer research, are considered for publication, with the understanding that they have not been published previously and are submitted exclusively to the *Journal of the National Cancer Institute Monographs.* All material submitted for consideration will be subject to review, when appropriate, by at least one outside reviewer and one member of the Editorial Board of the *Journal of the National Cancer Institute.* Opinions expressed by the authors are not necessarily those of the publisher or the editors.

Proposals for monographs should be submitted to the Editor-in-Chief, *Journal of the National Cancer Institute,* Oxford University Press, P.O. Box 31110, Bethesda, MD 20824-1110.

Journal of the National Cancer Institute Monographs are available to members of the Information Associates Program of the International Cancer Information Center (one copy of each monograph per member, on request prior to publication). Monographs are also available from Oxford University Press, Journals Subscription Department, 2001 Evans Road, Cary, NC 27513 USA. Telephone toll-free in the U.S. and Canada, 1-800-852-7323, or 919-677-0977. Fax: 919-677-1714. E-mail: jnlorders@oup-usa.org.

© Oxford University Press

Integrating Economic Analysis Into Cancer Clinical Trials: the National Cancer Institute–American Society of Clinical Oncology Economics Workbook

Supported by an unrestricted educational grant from Amgen, Inc.

Writing Committee

Martin Brown, Ph.D.
National Cancer Institute
Bethesda, MD

Henry A. Glick, M.A.
University of Pennsylvania
Philadelphia, PA

Frank Harrell, Ph.D.
University of Virginia
Charlottesville, VA

James Herndon, Ph.D.
Cancer and Leukemia Group B Statistical Center
Durham, NC

Mary McCabe, R.N.
National Cancer Institute
Bethesda, MD

Carol Moinpour, Ph.D.
Southwest Oncology Group
Seattle, WA

Kevin A. Schulman, M.D., Chairperson
Georgetown University Medical Center
Washington, DC

Tom Smith, M.D.
Massey Cancer Center
Richmond, VA

Jane Weeks, M.D.
Dana Farber Cancer Institute
Boston, MA

Damon M. Seils, Contributing Editor
Georgetown University Medical Center
Washington, DC

Scientific Editors

Martin Brown, Ph.D.; Mary McCabe, R.N.; Kevin A. Schulman, M.D.

The NCI–ASCO Economics Working Group

Charles L. Bennett, M.D., Ph.D.; Martin Brown, Ph.D.; Haim Erder, Ph.D.; Richard D. Gelber, Ph.D.; Henry A. Glick, M.A.; Frank Harrell, Ph.D.; James Herndon, Ph.D.; Bruce E. Hillner, M.D.; John Hornberger, M.D., M.S.; Mary McCabe, R.N.; Carol Moinpour, Ph.D.; Bernie J. O'Brien, Ph.D.; Arnold Potosky, Ph.D.; Scott Ramsey, M.D., Ph.D.; Kevin A. Schulman, M.D., Chairperson; Jeffrey Silber, M.D., Ph.D.; Tom Smith, M.D.; Koen Torfs; Edward Trimble, M.D.; Richard Ungerleider, M.D.; Jane Weeks, M.D.; Richard Willke, Ph.D.; and Robert E. Wittes, M.D.

Integrating Economic Analysis Into Cancer Clinical Trials: the National Cancer Institute–American Society of Clinical Oncology Economics Workbook

1998
Number 24

Contents

INTRODUCTION 1

PART I: ECONOMIC ANALYSIS AND CANCER CLINICAL TRIALS 1

Why Economic Analysis? 1
- The Rising Costs of Cancer Care 1
- The Need to Improve Decision-Making 1
- Perspectives on the Use of Resources 3

When Are Economic Evaluations Justified? 3
- Significant Resources at Stake 3
- Resource Considerations Are Prominent 3
- Resource Allocation Decisions Are Imminent 3

When Not to Use Economic Analysis 3

Generalizability of Observed Costs of Cancer Care 3
- Direct Medical Costs 4
- Direct Nonmedical Costs 4
- Productivity Costs 4
- Intangible Costs 4

The Need to Make Decisions 4

Criteria for Evaluating the Soundness of an Economic Analysis 4

Conclusion 5

PART II: PLANNING AN ECONOMIC ANALYSIS 5

Selection of Appropriate Trials 5

Timing of Data Collection 5

Choice of Perspective 6

Time Horizon 6
- Outside the Time Horizon 6
- Sequelae 6
- Relating Costs and Benefits 6

Endpoints 6

Baseline Characteristics 7

External Validity 7

Salvage Protocols 7

Importance of Early Collaboration Among Investigators 7

Pilot Testing 8

Conclusion 8

PART III: IMPLEMENTING AN ECONOMIC ANALYSIS 8

Study Outline 8
Resource-Use Categories 8
Direct Medical Costs: Inpatient and Outpatient Resource Utilization 9
Hospitalizations 9
Physician Services 9
Surgical Procedures 9
Diagnostic Tests 9
Pharmaceutical Services 9
Radiation Oncology 9
Ancillary Services 9
Subacute-Care Facilities 9
Direct Nonmedical Costs 10
Measuring Resource Utilization 10
Medical Records 10
Administrative Data Sources 10
Patient Self-Report 10
Assigning Unit Costs: Direct Medical Costs 10
Hospital Costs 11
Physician Costs 11
Nursing Services 12
Diagnostic Tests 12
Pharmaceutical Services 12
Subacute-Care Facilities 12
Capital Costs 12
Protocol Costs 12
Assigning Unit Costs: Direct Nonmedical Costs 12
Aggregating Price and Quantity 12
Generic Economic Evaluation Strategies 12
Inpatient Studies 12
Outpatient Studies 13
Subacute-Care Studies 13
Conclusion 13

PART IV: DESIGNING AN EFFECTIVE DATA COLLECTION STRATEGY 13

Designing Data Collection Forms 13
Hospital Resource Data 13
Outpatient Resource Data 13
Subacute-Care Institutional Resource Use 13
Caregiver Support 13
Covariates 14
Collecting Data 14
Beginning Data Collection 14
Why Collect Nonstudy-Site Data? 14
Obtaining Informed Consent 14
Maintaining Communication Between Research Teams 14
Conclusion 14

PART V: STATISTICAL ISSUES 15

The Importance of a Statistical Analysis Plan 15
Important Properties of Economic Data 15
Anticipated Uses of Economic Data 15
The Complexity of Economic Data 15
Distributional Characteristics of Economic Data 16
Statistical Methods for Analyzing Economic Data 16
Descriptive Methods 16
Univariable Methods 16
Covariable Adjustment and Multivariable Modeling Methods 17
Analysis in the Presence of Informative Censoring 18
Survival Techniques 18
Missing Data 18
Item Nonresponse 19
Unit Nonresponse 19
Imputation Methods for Missing Data 19
Variance of Price and Quantity Vectors 19
Statistical Issues Related to Cost-Effectiveness Ratios 20
What Should Be in the Denominator? 20
Variance of the Cost-Effectiveness Ratio 20
Sample-Size Analysis 21
Price Adjustments 21
Discounting the Costs and Benefits of Cancer Care 22
Conclusion 22

PART VI: ECONOMIC EVALUATION IN THE COOPERATIVE GROUP SETTING 22

Are the Data Required? 22
Who Will Collect the Data at Cooperative Group Institutions? 22
How Will Data Be Collected, Monitored, and Submitted? 22
Training: An Important Quality-Control Procedure 23
Small-Scale Pilot Studies 23
Who Will Analyze and Share the Data? 23
Sharing of Data 23
Data Analysis Issues 24
Intergroup Collaboration Issues Specific to Cost Studies 24
Quality Control 24
Intergroup Pilot Test 24
Data Issues 24
Information-Sharing 24
How Will Economic Studies Be Funded in Cooperative Group Trials? 24

CONCLUSION 26

REFERENCES 26

NOTES 28

Integrating Economic Analysis Into Cancer Clinical Trials: the National Cancer Institute–American Society of Clinical Oncology Economics Workbook

Introduction

Clinical economics is a new and evolving discipline that addresses the economic implications of changes in medical practice. As applied to cancer care, clinical economics assesses the costs and effectiveness of new cancer interventions and can be a valuable endpoint in selected clinical trials. Through the integration of economics into clinical evaluations, information can be developed that contributes to the decisions of patients, clinicians, health care managers, and policymakers as to the most effective allocation of cancer care resources.

To begin a formal effort to promote the development of economic analyses in National Cancer Institute (NCI) clinical trials, NCI sponsored a conference in 1994 with cancer center and cooperative group representatives to initiate discussions on the importance, appropriateness, and complexity of such evaluations. In 1995, the American Society of Clinical Oncology (ASCO) established its Health Economics Working Group with a charge to develop specific guidelines for implementing economic evaluations in cancer clinical trials. As a follow-up to these initiatives, in 1996, NCI and ASCO convened a workshop to consider the practical implementation of economic evaluations in cancer clinical trials. The participants in this small workshop included experts from cooperative groups, NCI staff, and other experts who are actively involved in health economics.

This workbook is the product of the meeting and subsequent discussions by the participants. It is meant to identify and elucidate the important characteristics of economic studies in the context of clinical trials, to indicate the considerations that investigators should address in their planning and implementation of such studies, and to suggest possible approaches. The workbook is neither a definitive text defining how all aspects of such studies should be handled nor an official NCI document prescribing how studies must be done. Rather, it is a developing guide to be used as a practical reference that will be revised as the state of the art progresses. The writing committee hopes that the workbook will serve as a useful tool for NCI cooperative groups as they incorporate economics as a research endpoint into the evaluation of new cancer treatment, prevention, and diagnosis strategies.

Part I: Economic Analysis and Cancer Clinical Trials

This section briefly presents the framework of economic theory underlying cost-effectiveness analysis and its application to the field of oncology. It is not, however, an in-depth review of the theory of economic evaluation, because many comprehensive sources currently exist *(1,2)*.

Why Economic Analysis?

The Rising Costs of Cancer Care

The percentage of total deaths in the United States attributable to cancer has risen from 16.3% in 1965 to 23.3% in 1997 *(3)*. From 1990 through 1996, the estimated costs of cancer treatment increased from $35 billion *(4)* to $50 billion due to higher inflation, increasing numbers of procedures and cases, and the aging of the population *(5)*. Even conservative estimates, measuring only the direct costs of treatment, show cost increases from $18.1 billion in 1985 to $27.5 billion in 1990 to $41.4 billion in 1994 *(6)*. Cancer will become the foremost cause of death in the United States by the end of the decade.

Statistics and predictions such as these underscore the likelihood that cancer will continue to absorb more of the increasingly limited resources of the U.S. health care system. Some researchers have suggested that increasing costs and demands from a sophisticated patient audience will require both implicit and explicit rationing *(7,8)*. In any event, economic forces will contribute to a growing need to better evaluate treatment practices by all clinicians, thus insuring that we utilize the relatively scarce resources of the health care system in an appropriate manner.

Cancer therapies are increasingly resource intensive, as evidenced by stem cell transplantation for hematologic disorders and solid tumors, paclitaxel for palliative chemotherapy of breast and ovarian cancers as well as non-small-cell lung cancer *(9)*, serotonin–antagonist antiemetics, and growth factors for supportive care during treatment. In an era of capitated physician and hospital payments, the resources available for cancer treatment will be increasingly constrained, and payers and purchasers will want to understand the value of cancer treatments, especially for resource-intensive therapies, before allowing widespread access to them. Furthermore, when health outcomes are identical for alternative therapies, the costs of these treatments may be the most important factor in determining whether to recommend or reimburse one of the treatments.

Encouraging examples of less intensive strategies have the potential to improve efficiency of cancer treatment, such as evidence-based, minimalist follow-up care for breast cancer patients *(10)*, decreased use of tumor markers in breast and colorectal cancers *(11)*, and conservative use of hematopoietic growth factors *(12)*, as well as a shift to less expensive outpatient treatments for stem cell transplantation *(13)*. However, these strategies remain a minority of cases of new treatments in oncology.

The Need to Improve Decision-Making

Because few medical interventions that provide health benefits also result in reduced health care expenditures (Table 1),

Table 1. Efficacy and cost-effectiveness of selected cancer and other treatments*

Intervention	Benefit	Cost/life-year (U.S.$ 1992)
IV IgG for CLL	0.8 days	+7 100 000
CEA in monitoring colon cancer	0.1% increase in OS	31 000 to 6 599 000
Aggressive Dx and Rx of unknown primary cancer	11%–11.5% increase in OS	2 400 000
ABMT for limited metastatic breast compared with standard treatment	6.0 mo	116 000
Captopril for moderate hypertension		82 600
Zidovudine for HIV	May improve survival	8200 to 88 500
Screening proctosigmoidoscopy for colon cancer		43 000 to 47 000
Adjuvant CMF for breast cancer in 75-year-old woman	1.8 mo	44 000
Renal dialysis, in-center benefit, men		42 000 to 80 300
Screening mammography, women >50 y	Reduces mortality by 31%	20 000 to 50 000
TPA vs. streptokinase for acute myocardial infarction	1/100 patients alive at 2 y	31 000
ABMT for Hodgkin's disease, second relapse only	10.3 y	26 200
Hydrochlorothiazide for moderate hypertension		23 500
Palliative chemotherapy for GI cancers	5 mo	20 200
Coronary artery bypass, left main disease + angina	Reduces deaths	17 400
Chemotherapy vs. AlloBMT for ANLL	OS 21%–48%	12 500
Chemotherapy for non-small-cell lung cancer	7–15 weeks	−8400 to +20 000
Adjuvant CMF for breast cancer	9.2–10.7 mo	4900
Adjuvant chemotherapy for Dukes' C colon cancer	≥1 y	2000 to 3000
Smoking cessation counseling, men		1300
Cost-saving interventions		All<0
Total hip replacement		
Pneumovax		
Influenza vaccine		
Hepatitis B vaccine		
Screening for familial hemachromatosis		
Coordinated home care for terminally ill cancer patients		

*Modified from Smith et al. *(94)*. IV = intravenous; IgG = immunoglobulin; CLL = chronic lymphocytic leukemia; CEA = carcinoembryonic antigen; Dx = diagnosis; Rx = treatment; OS = overall survival; ABMT = autologous bone marrow transplantation; HIV = human immunodeficiency virus; CMF = cytoxan, methotrexate, 5-flourouracil; TPA = tissue plasminogen activator; GI = gastrointestinal; AlloBMT = allogeneic bone marrow transplantation; and ANLL = acute nonlymphocytic leukemia.

reliable economic analyses must be developed to improve our decision-making about the allocation of resources for cancer treatments *(9,14,15)*. The rapid advancement of science and biotechnology, coupled with increasing fiscal pressures, may severely limit access to cancer treatments for which economic data are not available. It is worth remembering that medical ethicists have argued that the least ethical way to allocate resources is to continue to spend money on unexamined treatments *(16,17)*.

As shown in Table 1, most treatments provide additional clinical benefits at an additional cost. Thus, economic efficiency does not imply a reduction in cost of care but provides a measure of the value provided for the cost of the treatment *(18)*.

Economic information can strengthen arguments for adoption of new technologies that are economically attractive, provide additional input for decision-making in indifferent or gray areas, and provide clear-cut answers when costs are a major consideration *(19)*. Most existing cancer treatments have not been evaluated with the stringency applied to newer treatments. For example, we have no ''proof'' that intensive treatment of metastatic breast cancer, compared with best supportive care, improves quality or quantity of life or saves money, because no randomized clinical trial has been performed in the past 20 years. This is not to say that intensive treatments are ineffective or exceed reasonable cost-effectiveness limits—only that they have not been studied with the same scrutiny as high-dose chemotherapy with stem cell transplantation *(20)*.

Measurement of incremental benefit (additional lives saved by a strategy) and incremental cost (additional costs of a strategy) forms the backbone of comparative economic analysis. The four possible approaches of such analysis are listed in Table 2 and represent the combinations of costs and outcomes measurement. Estimates of absolute incremental differences in cost between trial arms are needed, just as they are for analyses of

Table 2. Potential combinations of costs and outcomes

Costs and outcomes	Results	Example
Improved outcome and decreased costs	A strategy with improved outcome and lower costs is dominant and would always be adopted.	Peripheral blood stem cells mobilized by hematopoietic growth factors allow earlier hospital discharge and offer cost savings of about 20% *(95)*.
Improved outcome and increased costs	Clinical advantage at an increased cost.	Paclitaxel instead of cyclophosphamide and cisplatin for advanced ovarian cancer results in improved survival at an additional cost of $20 000 to $30 000 per life-year *(96)*.
Poorer outcome and decreased costs	Decrease in clinical advantage but at a reduced cost.	Generic prochlorperazine instead of ondansetron for emesis less effective, but drug costs are markedly less.
Poorer outcome and increased costs	Poorer outcome and increased costs should not be adopted.	Mitoxantrone instead of doxorubicin in treatment of metastatic breast cancer does not improve survival or quality of life and has higher costs.

treatment benefits. For example, the primary outcome measure of a cost-effectiveness analysis is the additional cost incurred per year of life gained or quality-adjusted years of life gained. Such measurement requires explicit delineation of clinical benefit in terms of years of life saved or quality-adjusted years of life saved and of dollars spent on treatment, recurrence, survival, and death. The statistical implications of absolute and relative differences are discussed in Part V.

Perspectives on the Use of Resources

Cost data are meaningful only when considered from a specific perspective, usually that of patient, caregiver, provider, payer, or society (i.e., the choice of perspective determines what events or expenditures are regarded as costs, as well as how those costs should be calculated). The societal perspective is often the most appropriate one for resource-allocation decisions *(1)*, because its broad approach can minimize bias in comparisons; however, other perspectives may be important in attempts to understand the economic effects of a treatment.

As in clinical medicine, investigators must not make assumptions about how patients value treatment outcomes. For example, a recent clinical trial demonstrated substantial intellectual deficits in glioblastoma patients undergoing radiotherapy; but, because the treatment offered a few weeks of improved survival, most patients reported that they would choose the radiotherapy *(21,22)*.

Multiple perspectives may be adopted within a single analysis to illustrate the views of patients, payers, policymakers, and health care providers. For example, the cost of providing a hospital service is usually different than the charge for the service and the amount paid by the insurance company. An episode of care may cost a hospital $7000, but the hospital charges $10 000 for the service. From the hospital's perspective, the charge is an overstatement of the resources consumed. If the patient must pay the full charge, however, that charge accurately represents cost from the patient's perspective. Alternatively, while a hospital can sometimes decrease its costs by discharging patients early, patients' costs may increase due to increased outpatient expenses that are not covered by health insurance.

When Are Economic Evaluations Justified?

Inclusion of economic evaluations in clinical trials can be justified in a variety of circumstances: when significant resources are at stake; when resource considerations are prominent; and when resource allocation decisions are imminent *(23,24)*. Examples of each circumstance are provided below.

Significant Resources at Stake

Antiemetics. Although individual doses of 5HT3-agonists, such as odansetron and granisetron, are not expensive, they carry a high potential volume of use and can account for 5% of a hospital's oncology pharmacy budget (data on file, Medical College of Virginia Hospitals).

Colony-stimulating factors (CSFs). The initial and additional costs of CSFs (granulocyte, granulocyte-macrophage, and erythropoietin) can consume up to 5% of a hospital's oncology pharmacy budget (data on file, Medical College of Virginia Hospitals).

Transplant procedures. High-dose chemotherapy with stem cell transplantation and similar experimental therapies carry a high dollar cost, regardless of volume. The cost of treatment and follow-up for leukemia and lymphoma patients receiving autologous bone marrow transplantations and granulocyte–macrophage CSF reached $79 892 *(25)*.

Genetic predictors. BRCA1 and BRCA2 tests, genetic predictors for breast and ovarian cancer, can cost anywhere from $150 to $1500 per mutation, depending on whether a family's specific mutation has been identified. Lerman et al. *(26)* note that the high cost of such testing may deter some individuals from undergoing the test, and individuals may hesitate to request coverage from insurers for fear of future discrimination.

Resource Considerations Are Prominent

Many therapies carry a high cost, regardless of their clinical outcome, such as high-dose chemotherapy with stem-cell transplantation for metastatic breast cancer and allogeneic bone marrow or stem cell transplantation from an unrelated matched donor (estimates allow $100 000–$250 000 for each uncomplicated case).

Other therapies vary widely in costs but produce similar clinical outcomes. For example, high-dose chemotherapy is commonly used as treatment for first-remission lymphoma, although less expensive standard-dose chemotherapy is known to produce similar or identical outcomes *(27)*.

Resource Allocation Decisions Are Imminent

Megakaryocyte growth factors have been proposed to eliminate the need for platelet transfusions and to allow for earlier hospital discharge after high-dose chemotherapy *(25)*. Also, results of a trial of high-dose chemotherapy as adjuvant therapy for high-risk breast cancer (>10 positive lymph nodes) quickly changed the standard therapeutic approach to this disease *(28)*.

When Not to Use Economic Analysis

Economic analysis is unnecessary when a primary therapy works well in a small number of patients—for example, testicular cancer treated with platinum (600 cases/year) *(29)*. Furthermore, some therapies vary little in costs and produce similar clinical outcomes in very common diseases. For example, a number of chemotherapeutic combinations are used as adjuvant therapy for node-positive breast cancer, but they cost about the same and produce similar effects *(30)*.

Also, economic analysis is unnecessary when its results are unlikely to alter clinical practice. For example, an economic analysis of breast self-examination might show that costs are higher in the screened population and fail to demonstrate clinical benefit. But the use of breast self-examination is so widespread that an economic analysis would not change the practice.

Generalizability of Observed Costs of Cancer Care

Cost reflects the resources needed to provide a service and it has two elements: "resource utilization" in some natural units (e.g., one hospital day or one dose of ceftazadime) and the "unit cost" or price (e.g., 1 hospital day in a semiprivate room for $460). Units and prices are multiplied to determine treatment costs. One might calculate physician costs by multiplying physician time by hourly physician earnings. The same principle

applies to productivity costs: the cost of patient time can be calculated by multiplying time by a patient's daily wage rate.

Some administrative datasets provide reliable information only on resource utilization, whereas others provide reliable information only on total cost, and still others provide information on both. It is important to recognize that the resource utilization and price components of cost have different determinants. For example, resource utilization may be affected by treatment setting, organizational and technical innovation, and efficient use of fixed resources. Price may be affected by the degree of competition in health care markets, including geographic location, regulatory and legal restraints, and the bargaining power of health care payer organizations. The "learning curve" of better efficiency and lower cost that develops as procedures are performed more often may decrease some costs; alternatively, unforeseen consequences may increase expected costs. It is important to keep these issues in mind, because economic data obtained from one clinical trial may not always be generalizable to another time, setting, or pricing scheme.

Costs related to a disease entity have traditionally been classified into a number of categories, depending on whether the costs are directly related to the expenditure of medical or nonmedical resources related to treatment and care; whether the costs are a reflection of lost economic productivity due to disease-related disability or premature mortality; or whether the costs are related to the pain and suffering caused by the disease *(6)*. The four categories of costs are listed below:

Direct Medical Costs

- Resource use related to services and procedures delivered by the health care system in the course of treatment and care (hospital and physician costs); patient payment for medical supplies.

Direct Nonmedical Costs

- Resources expended by the patient and her family related to treatment and care (travel to treatment facility, parking, etc.).

Productivity Costs

- Not a financial transaction, but the monetary value of lost economic production due to morbidity or mortality, often utilized in cost-effectiveness or cost-utility analyses; may be important from a societal perspective or from the perspective of an employer—for example, in the analysis of treatments for substance abuse or treatments that effect the ability of cancer patients to return to work.

Intangible Costs

- Also not a financial transaction, but the monetary expression of pain and suffering associated with disease or treatment; various methods for measuring intangible costs, such as "willingness to pay," indicate that such "costs" can be quantified as a monetary value (i.e., the amount a patient would be willing to pay to avoid pain or suffering); for purposes of cost-benefit analysis, intangible costs are typically expressed in monetary terms, because the outcome of cost-benefit analysis is expressed in purely monetary terms; for cost-effectiveness analysis, intangible costs often are ignored, and for purposes of cost-utility analysis, intangible costs are expressed as part of the denominator in units of quality-adjusted life-years (QALYs)

While all of the categories of costs are included in cost-of-illness studies and cost-benefit analyses, only direct costs are usually included in cost-effectiveness (or cost-utility) analyses, which have become the most common approach in economic analysis of health care interventions (Table 3).

The Need to Make Decisions

The need for either implicit or explicit decision-making about resource allocation dictates that decision makers should know the total budget of a program, the efficiency or cost-effectiveness of the given modality, and the overall effect of each treatment. It is essential, therefore, that data be presented in such a way that decision makers can assess both the costs and potential clinical consequences of new cancer treatments (Table 3). Although the mechanics of such decision-making vary greatly from centralized planning to individual choice, only limited economic information is currently available to make these important treatment decisions.

Criteria for Evaluating the Soundness of an Economic Analysis

Economic analyses must be evaluated using defined criteria with clear goals in mind, just as clinical trials themselves are evaluated against a variety of accepted criteria. A full economic analysis of any cancer intervention should go beyond a simple

Table 3. Outcome and cost studies for presentation to decision makers

Type of study	Attributes examined	Comments
Clinical outcomes	Clinical consequences only.	Ignores costs; easy to choose among clearly superior therapies such as cisplatin for testicular cancer; more difficult among all others that give lesser benefits at high costs.
Cost analysis	Costs only, e.g., costs of treating febrile neutropenia.	Ignores clinical outcomes; does not help decision makers choosing among clinical strategies unless clinical outcomes are identical.
Cost-effectiveness	Measures additional benefit of one strategy over another in years; measures additional costs of that strategy in dollars; combines to give dollars per year of life gained ($/life-year).	Most appropriate for comparisons of alternative therapies to treat the same condition. For example, different chemotherapy regimens for small-cell lung cancer. Cannot be used to compare treatments across disease states.
Cost-utility	Adjusts benefit gained by a utility value, or value given to being in a certain health state.	Most appropriate for comparisons of alternative therapies to treat the same condition as well as for general comparisons of interventions across disease states.

identification of costs (Table 4). As the competition for scarce health care dollars intensifies, more therapies will be marketed as "cost-effective." Failure to explain these criteria should make the investigator wary about the quality of the economic analysis *(31)*.

Conclusion

In this section, we review the rationale for economic evaluation of new cancer treatments, including a discussion of why economic analysis is increasingly being implemented in cancer clinical trials, a discussion of when economic evaluations are justified, and an overview of the economic framework for evaluation of cancer clinical therapies.

Economic evaluation is an increasingly important component of clinical decision-making in the context of scarce or limited resources. Economic evaluation itself can contribute to decision-making in specific clinical settings as outlined in this section. Furthermore, economic evaluation goes beyond assessing the monetary cost of medical care to include such issues as productivity costs (loss to the economy from morbidity and mortality) and intangible costs (costs of pain and suffering related to illness or injury).

Part II: Planning an Economic Analysis

The previous section introduced some of the concepts of economic evaluation of new cancer therapies by reviewing the conceptual framework underlying economic evaluation of cancer clinical trials. Parts III through V will review these concepts in more detail and lay out a framework for the collection of economic data in clinical trials. The purpose of this section is to illuminate issues in the planning of economic analyses that may be important to clinical investigators designing protocols with an economic component.

Table 4. Criteria for identifying a sound economic analysis*

Criterion	Comment
Competing alternative strategies described	"Cost-effective" compared to what specific alternative?
Incremental costs and clinical consequences described	Should be a cost/life-year gained, "apples-to-apples" comparison
Sensitivity analysis of differing assumptions performed	Costs and clinical outcomes may differ, so all assumptions should be justified
Costs and consequences discounted	Time and dollars in the present worth more than in the future
All important costs and consequences identified for each strategy	Should include all relevant costs, e.g., both inpatient and outpatient
Well-defined question	Was the cost study part of the initial design or done afterwards?
Costs and consequences valued credibly, in usual terms	Should be stated in $/life-year; watch out for ratios made up just for a particular study
Costs and consequences measured accurately in understandable units	Should be in $/life-year or similar format
Published in a peer-reviewed journal	Given market pressures, more and more such articles will show up in industry-sponsored journal supplements without the rigor of peer-review

*Adapted from *(2)*.

Selection of Appropriate Trials

Economic evaluation can be an important secondary endpoint in clinical trials of new drugs and technologies, especially new cancer therapies. Implementing an economic evaluation, however, requires additional time and effort, including additional data collection, management, and analysis. In deciding whether to make this investment, investigators must consider whether economic analysis is likely to aid in decision-making regarding reimbursement for new therapies or technologies.

In general, investigators have suggested that economic evaluation is most appropriate in pivotal trials of new technologies or strategies when the investigational therapy is resource intensive or is likely to be utilized by a significant number of patients. Economic evaluation is less important in trials of "me-too" therapies, which may be similar in efficacy and cost to other clinical therapies. Here, however, the cost of therapy may be the important distinction between two treatment arms. Clinical investigators should develop economic hypotheses before approaching economic investigators to discuss the inclusion of economic analysis as an endpoint in the clinical trial.

Economic evaluations usually are proposed for inclusion in phase III clinical trials but can be included in phase II trials to allow for collection of pilot data that will be used in plans of pivotal phase III assessments. Phase IV studies are also appropriate for consideration for economic evaluation.

Some estimate of the potential difference in cost between treatment arms, and the resulting cost-effectiveness ratio if the targeted difference in survival is achieved, should be included in the concept sheet and protocol for any study with an economic component. Ideally, this estimate will be based on data collected in a pilot study, but a "back-of-the-envelope" calculation should be performed when pilot data are not available. Such estimates play a critical role in justifying the investment in an economic component and are needed for the planning of an efficient data collection strategy.

Timing of Data Collection

Data collection for economic studies can be based on prospective data that have been collected concurrently with the clinical trial or on retrospective data that have been collected after the trial's completion. The advantages and disadvantages of each approach are discussed in more detail in Part V.

Prospective economic evaluation assures that economic data will be available at the time of a trial's completion. It allows for collection of the four types of economic data (direct medical, direct nonmedical, productivity, and intangible), as well as quality-of-life data from study patients. But prospective collection of economic data requires substantial planning and is, in this sense, no different than prospective collection of clinical data. Because economic evaluation is usually a secondary endpoint of clinical trials, and because there is a risk that the null hypothesis will be adopted at the end of the trial, investigators should carefully consider the potential use of resources for an economic evaluation in a trial of an unknown or untested therapy.

Retrospective economic evaluation offers a more limited approach to assessment of resource use within clinical protocols,

typically focusing on direct medical costs, because other types of data often are not available in the medical record. In addition, retrospective approaches can not assess patients' preferences or quality of life. However, in this approach, the expense of collecting economic data will only be required for therapies with proven efficacy.

The choice between prospective and retrospective data collection usually rests on the potential clinical effect of the new therapy and on the amount of resources available to investigators. If the trial is indeed an appropriate candidate for economic evaluation, and if resources are available for proper planning and design of an economic study, a prospective approach should be taken, because it will allow for a more thorough evaluation of the costs of a new therapy or technology.

Choice of Perspective

Economic evaluation of medical care is unique in that it assesses the costs and benefits of therapies from a variety of perspectives, including those of patients, caregivers, employers, payers, and society. However, clinical and economic investigators must determine which perspectives are most relevant to the clinical trial under consideration. For example, economic evaluation from a hospital's perspective requires consideration of inpatient costs but not outpatient or out-of-pocket costs. Data collection from a patient's perspective is more complex and may require primary data collection to yield more useful information on out-of-pocket costs and lost work days. Economic evaluation from a societal perspective is the most comprehensive of all and includes consideration of both patient and provider perspectives.

Unless a therapy is particularly burdensome to patients, the most common perspective for clinical trials is a "modified" societal perspective, which considers only direct medical costs. In this type of analysis, direct medical costs are quantified as social costs, but direct nonmedical and productivity costs are not routinely collected.

Time Horizon

Time horizon is defined as the time from randomization to follow-up for patients in clinical trials. Time horizon can affect the economic evaluation of new therapies by limiting the amount of information available to the assessment of the long-term effectiveness of treatment. Economic evaluation frequently requires a different time horizon than that proposed by clinical investigators.

Ideally, the time horizon of an economic evaluation should be such that all costs related to the therapies being compared are captured for analysis. This concept differs from the idea of time horizon in clinical trials, which can be based on time until a measurable clinical endpoint is achieved or until the end of some predetermined time period (e.g., 5 years). Under the ideal economic scenario, subjects would be randomized to a trial and followed until death or cure *(32)*. Unfortunately, most clinical trials are conducted for a limited length of time.

In general, the time horizon for an economic evaluation of a clinical trial begins at the date of trial randomization. Alternative start dates for the time horizon include time of diagnosis or time of the onset of symptoms. Economic evaluations that begin before the time of randomization usually are attempting to assess important costs incurred during the prerandomization time period (e.g., a trial of bone marrow transplantation for patients with breast cancer may need to provide induction chemotherapy to 10 patients to find one responder for the transplant portion of the study; an investigator may be interested in the costs of screening as well as the costs of the prerandomization hospitalization) *(33)*.

In determining the appropriate length of the time horizon for economic evaluation, it is helpful to distinguish between two types of clinical trials—trials of therapies for metastatic cancer (mortality studies) and trials of adjuvant therapies. Clinical trials in these two settings pose different issues with respect to clinical and economic endpoints. Knowledge of the survival curves for the type of cancer being studied, along with projections of the expected benefit of the therapy, can be used to determine how long patients should be followed for an economic study. In mortality studies, the follow-up time is likely to be relatively short, so resource utilization can be assessed for all patients until time of death for a high proportion of study patients. In clinical trials of adjuvant therapies, some proportion of patients may be cured and thus live a normal life expectancy. Generally, patients can be followed for recurrence. In adjuvant therapy trials, one can consider three potential time frames—short-term (e.g., 6 months, 1 year), intermediate, and lifetime.

Outside the Time Horizon

When the time horizon of a clinical study differs from that of the economic study, it may be necessary to track all resource use in both arms of the trial over a period that is longer than the relevant clinical endpoint, because important resource use may occur in the period beyond the time horizon of the clinical trial.

Sequelae

Substantial long-term sequelae that may occur in the relatively distant future, whether related to the cancer itself or to treatment of the cancer, should be included in the analysis.

Relating Costs and Benefits

The time horizon should be the same for the numerator and denominator of a cost-effectiveness ratio. For example, if an analysis includes 2 years of clinical follow-up data, it should also include 2 years of cost data. In a trial of adjuvant therapy, simply including the cost of treatment in the numerator and a longer follow-up in the denominator does not accurately reflect the relationship between the costs and benefits of the therapy being assessed.

Endpoints

Economic evaluation usually is based on a consideration of final outcomes, such as survival rather than response, disease-free survival, and time to progression. Thus, economic evaluation requires either consideration of a final outcome measure as an endpoint of the clinical trial or an understanding of how clinical trial endpoints will translate into final endpoints (for example, epidemiologic data may be available for estimations of cure rates for patients who demonstrate a complete response to treatment). Specification of outcomes measures must be determined prospectively by both clinical and economic investigators before the trial proceeds.

Baseline Characteristics

Clinical studies, except for some "large and simple" trials and a small number of special cases, are generally underpowered for evaluating secondary cost or quality-of-life measurements. Analytic techniques that assess nontreatment-related variance can be used to address this issue. Thus, the ability to detect differences in costs due to treatment assignment will increase if information on baseline characteristics of study patients is included as a set of covariates in analyses of secondary study endpoints. Baseline characteristics include resource use in the period prior to the trial, clinical status, diagnostic profile, disease stage and severity, level of social support, and health status. Variation in covariates among patients stem from the type of patient rather than from treatment assignment—for example, whether a patient is an inpatient or outpatient at the time of randomization can predict resource utilization after randomization.

Baseline covariates allow investigators to "disentangle" patient characteristics prospectively. Covariates may be used as predictors in regression analysis at the end of a study, or they can be used to stratify the randomization of any parameter that may be a factor in predicting costs to a high degree. Many of these data elements can be collected as part of eligibility determination and tagged as covariates in the economic evaluation. It is also important to consider any and all clinical information obtained prior to randomization as potential covariates. Ideally, covariates will be continuous variables.

External Validity

Where possible, the clinical protocol should be modified to reduce protocol-mandated tests or procedures to reduce protocol-induced charges in medical care (protocol-induced costs or benefits) and to ensure that the protocol mirrors "usual care" as much as possible. Economic investigators should also review the clinical protocol to ensure that there are no economic biases in the structure of the trial (e.g., there are no fixed discharge criteria included in the study and/or there are no differences in prescribed treatment across study areas). In addition, whether or not the clinical protocol procedures are modified to reduce mandated tests, efforts should be made to identify medical procedures that are mandated only for the research portion of the study. If these research resources are not related to treatment, their costs can be excluded from the analysis.

Economic evaluation raises a unique set of issues about the generalizability of clinical trials themselves. These studies are designed to influence clinicians, patients, and decision makers in their perceptions of the importance of a new therapy. Thus, economic evaluation aims to ensure that patients being studied in the clinical trial are as similar as possible to relevant patient populations in the "real world." For appropriate economic evaluation, therefore, clinical trials should be free of economic bias, should include patient entry criteria broad enough to provide support for the external validity of the analysis, and should, to the greatest extent possible, follow strict and objective standards in patient selection.

Trial investigators can reduce economic bias in the design of clinical studies by avoiding prespecified indications for discharge, supportive care interventions, or second-line therapies. Investigators can further reduce bias by identifying and minimizing protocol-induced costs or benefits, thus rendering the phase III clinical trial as realistic as possible in terms of actual practice in the community. Protocol-induced costs can include repeated diagnostic tests or other evaluations that do not directly contribute to clinical care or are not routine elements of clinical practice. Because these tests and evaluations can lead to diagnoses that otherwise might be missed in the usual care of patients, they can also yield clinical outcomes that would not be replicated in usual practice.

Finally, economic analysis is concerned with the generalizability of data developed from clinical trials. The external validity of studies is thus an important consideration in recruitment of subjects to the study. One such example of data that suggest clinical trial results can be replicated in the community are presented in a study of breast cancer mortality in British Columbia by Olivotto et al., showing that after adoption of routine adjuvant chemotherapy, mortality declined with an effect size identical to that of clinical trials *(34)*.

An alternative view is that patient selection will probably continue to be a strong factor in most trials. The task for researchers is to determine how well bias to external validity can be modeled by obtaining good data on the relevant characteristics of patients enrolled in trials compared to patients who are likely to receive the treatment shown to be efficacious in the trial.

Salvage Protocols

In oncology, the potential for salvage or late use of interventions is important to consider in advance for economic evaluation. As mentioned above, the economic period of a treatment may be in discordance with clinical trial measures. Late use of interventions may provide crucial information for economic evaluation of a study, because treatment for disease progression can be very costly.

Economic evaluation can be implemented in this setting by including patients in the clinical trial for a fixed time period, allowing patients to remain in the economic protocol throughout this period even if they reach an endpoint of the clinical trial and withdraw from active treatment. Salvage protocols with economic arms also provide an opportunity for further economic data collection from patients who have failed treatment.

Finally, several unique statistical issues play a role in economic evaluation. Patients may incur fairly low costs until their disease has progressed, after which time they usually are withdrawn from the clinical trial. Yet, it is often at this point that patients become expensive and, thus, important to the economic evaluation. Economic studies often are designed to allow patients the option of continuing with the economic protocol, even after withdrawing from active treatment on the clinical protocol. In this limited way, many economic evaluations have been able to successfully follow large portions of a trial population beyond the time horizon of the clinical trial *(25)*.

Importance of Early Collaboration Among Investigators

To be successful, economic investigators must understand the clinical aspects of the trial being considered as well as the potential clinical implications of the therapy or technology being

evaluated. Likewise, clinical investigators must be able to develop economic hypotheses and to understand all elements of the economic study to ensure appropriate implementation of the economic study's design. Thus, clinical–economic evaluation is a truly collaborative undertaking between clinical and economic investigators.

Economic investigators should become familiar with the economic profile of the intervention under study: What resources are required to implement the therapy? For how long will these resources be required? What is the potential benefit of therapy? What are the potential resource offsets or savings from the new therapy compared with existing therapies? In addition, economic investigators should ask these questions in the context of the current standard of care. They should consider the potential risks of the new therapy as well as how the clinical trial's results might change the standard of care.

Clinical investigators should understand that there is no "standard" economic analysis to be simply appended to a clinical trial. Economic analysis is an important secondary endpoint of a study and, to be performed well, must be integrated throughout the clinical trial mechanism. Seamless integration of the economic component of the study with the quality-of-life data collection strategy is particularly important in any study that employs patient diaries as instruments of economic data collection. Most important, clinical investigators must "buy in" to the economic evaluation so that they have a stake in seeing that economic data are treated with as much care as clinical data. Recipes for failure of an economic evaluation include lack of commitment by clinical investigators to the collection of economic data and a lack of understanding by clinical investigators of the purpose of the economic evaluation.

Pilot Testing

Economic evaluation can benefit from pilot testing in much the same way as clinical trials. Pilot testing offers investigators an opportunity to assess the use of case report forms (CRFs) and patient diaries in clinical trials so that they might become more familiar with the process of implementing a protocol and more knowledgeable about the economic profile of the disease under study. In addition, pilot tests can pinpoint a trial's "cost drivers" and help economic investigators characterize the distributions of cost data and predict the differences in cost between treatment arms or the total cost of care for study patients. Identifying cost drivers in a pilot study can reduce the data collection burden in the final study. Furthermore, pilot data can be helpful in identifying the toxicity profiles of a therapy and, in particular, the nature and frequency of high-cost events, such as hospitalizations for febrile neutropenia or cardiac toxicity. Similarly, pilot testing can provide information for estimations of sample size and expected cost differences between study arms *(35)*.

Conclusion

This section reviews issues important for clinical investigators to consider in contemplating whether to implement an economic evaluation within a specific clinical trial. Issues for clinical investigators to consider include the selection of appropriate clinical trials for economic evaluation, timing of data collection, choice of perspective, time horizon, study endpoints, and external validity or generalizability of the study. Investigators must also consider that to design robust economic comparisons of clinical trials, economic evaluation requires input from economic investigators as early collaborators within the clinical team. To implement a successful protocol, clinical investigators must assist in the design of the economic evaluation and in the implementation of the economic component of the protocol.

Part III: Implementing an Economic Analysis

In the previous section, we reviewed some of the important concepts for a clinical investigator to consider when deciding whether to incorporate an economic evaluation into a clinical trial. In this chapter, we review in more detail many of the same issues from the perspective of someone who is designing an economic evaluation of a clinical trial. In the next chapter, we focus on the collection of data items themselves.

Study Outline

Economic evaluation measures the use of resources and often takes note of the same events of interest to clinical investigators, such as whether a patient receives a magnetic resonance imaging scan (MRI). Yet, while clinical investigators are most interested in the results of the MRI, economic investigators take greater interest in the number of MRIs consumed during the study period. In addition, economic investigators attempt to interpret changes in resource quantities. For example, a specific therapy may reduce length of stay for patients receiving autologous bone marrow transplant (ABMT) yet increase the need for outpatient follow-up visits. Costs are applied to these resource-utilization measures to demonstrate overall economic benefit or harm. For example, if inpatient days cost $700 each and physician visits cost $100 each, the ABMT scenario described above will save money if it reduces length of stay by 1 day and increases outpatient resource use by no more than the equivalent of six physician visits.

Thus, economic evaluation requires both resource utilization and unit cost information to provide an overall estimate of the resources consumed in the course of caring for an individual patient. The following sections focus specifically on the development of resource-use data and unit costs.

Resource-Use Categories

Oncology patients consume a variety of resources in the course of medical treatment, including hospital services, physician and nursing services, diagnostic tests, pharmaceutical products, radiation therapy, and ancillary services. Given the tremendous complexity of oncology care, economic investigators must make explicit decisions about the types and amounts of information they wish to incorporate into their protocols. In general, these data elements should include resources that carry significant costs (e.g., inpatient stays) and services that are likely to differ across treatment groups. The most important of these—known as "cost drivers"—represent resources that are both costly and differ across treatment groups *(36)* and they should be validated through pilot data collection efforts. Cost drivers are not simply drivers of a study's total cost but of the increments in costs between the study's treatment arms. Investigators should, therefore, assess the impact of a particular resource on the mean costs, the variance in costs, and the difference in costs between

treatment arms before eliminating it from the list of data elements to be collected.

Computerized cost-accounting systems, which are increasingly being adopted by health care provider organizations, make it possible to retrieve information on resource use and costs for a variety of resource-use categories. For example, the Decision Support System used by Group Health Cooperative of Puget Sound contains information on resource use and costs for such categories as inpatient days, surgeries, outpatient/short-stay hospitalizations, primary care visits, specialty care visits, mental health visits, emergency room visits, community health services, diagnostic radiology, laboratory services, occupational/physical therapy, respiratory therapy, and pharmacy *(37)*.

Direct Medical Costs: Inpatient and Outpatient Resource Utilization

Hospitalizations

Hospitalizations typically represent the most intense period of resource utilization by clinical trial patients. Resources received during a hospitalization include hospital "hotel services" themselves, physician services, nursing services, laboratory and pathology services, blood work, operating rooms, and pharmaceutical products.

Hospital services are provided based on the type of unit to which the patient is assigned, e.g., intensive care, high care, telemetry, general, and rehabilitation. Thus, descriptions of resource utilization for hospital "hotel services" could be a description of length of stay or a description of length of stay by unit type (e.g., a 10-day admission of which 4 days were spent in the oncology unit and 6 were spent on the general medical floor). One distinguishing feature of the different unit types within a hospital is the level of nursing support services available to patients. Thus, location within the hospital is often also used as a description of the intensity of nursing care.

Physician Services

Physician services in the hospital setting include those provided by the attending physician (either medical or surgical) and by consulting physicians. In the outpatient setting, they can include oncology services as well as services provided by primary care and consulting physicians. In an economic investigation, measures of physician effort can be characterized in terms of time (a 30-minute consultation) or workload units (resource-based relative value units [RVUs]) *(38,39)*. These data can be collected for all physicians treating patients in the protocol or by the type of physician providing the service.

Surgical Procedures

Surgical procedures are an important component of physician services. Performed for diagnostic and therapeutic purposes, surgical procedures can occur in and out of the hospital setting. The most global measure of surgical services is a description of the procedure itself (e.g., laparoscopy, colectomy). This description can be text based, or it can be coded using a standardized system such as the International Classification of Diseases (ICD) or the Common Procedural Terminology (CPT). Surgical procedures can be further described using a second level of detail—for example, the time in the operating room or the use of specific supplies, including blood products or isotopes. A third measure, which is especially important for outpatient surgical services, is a description of the location where the service is performed (e.g., hospital, outpatient surgical center, free-standing surgical center, physician's office). These levels of detail can be used to describe resource utilization by different treatment modalities.

Diagnostic Tests

Some of the most important services provided to patients both in and out of the hospital are diagnostic tests. As mentioned above, economic investigators are interested in recording the number and types of services provided to patients. These resources can be captured in a variety of ways, including listing the number of radiologic or laboratory tests as either an aggregate measure or an itemized listing of the occurrence of specific types of tests and procedures.

Pharmaceutical Services

Pharmaceutical utilization during hospitalization periods may already be captured within clinical CRFs' concomitant medication sections. If not, investigators may wish to detail the amount of medication received by patients during a hospital stay. Medication use can be characterized by medication type and dose (total daily dose) or by total dose received during a specific time period (i.e., daily dose, total dose during a course of chemotherapy, total dose during a hospitalization). In addition, economic investigators should record the route of administration, because intravenous administration may be more costly than oral administration. Data collection can include all concomitant medications or a subset of specific medications of interest to the investigators.

Radiation Oncology

Physician services for radiation oncology treatment can be captured as described in the Physician Services section above. Use of radiation treatment facilities can be tracked by recording the types of services received by patients, the personnel effort required to provide such services, the duration of treatment, the dose of isotope delivered, and the setting where the treatment occurred. In radiation oncology protocols, more detailed treatment or service data may be collected.

Ancillary Services

Many other services are available to patients in and out of the hospital. These services can be especially important to the care of oncology patients, although they are not delivered by physicians. Social work, speech therapy, pain management, occupational therapy, physical therapy, and clinical pharmacology services, to name a few, are usually measured by tracking the occurrence of specific consultations or by measuring the amount of time the provider spent in consultation with study patients.

Subacute-Care Facilities

Cancer patients may utilize a variety of subacute-care facilities, including nursing homes, rehospitalization centers, and inpatient hospices. These resources may be quantified by the type of institution and the length of stay in the institution. For more specific studies, data can be collected on specific services received during the stay.

Direct Nonmedical Costs

Direct nonmedical costs include the costs of transportation to and from physician visits or hospitalizations and other costs related to receiving medical care. For extended stays at a treatment center, these costs can include the cost of hotel services or of moving families to a treatment center. Investigators should determine the specific resource quantities of interest to be captured as a direct nonmedical cost. These can include the number of hours of parking at specific treatment centers or the number of days spent at a hotel adjacent to a treatment center.

Measuring Resource Utilization

Data on resource measures can be collected from medical records and medical bills and through patient self-report. In this section we review the data sources available to economic investigators.

Medical Records

Medical records include source documents, such as medical charts and abstracted flow sheets. These documents may contain data on all elements of resource utilization during a specific time period, such as the detailed information found on a hospital chart, which can be abstracted by data collectors to provide information for each of the resource categories reported in the previous section. Interpretation may be required for some data elements, such as unit type by day of hospitalization. Medical records are limited in that they often are not linked between inpatient and outpatient treatment facilities. Patients typically have multiple providers and, thus, multiple records, making it difficult to follow patients from one site to another.

Administrative Data Sources

Many types of administrative data are available to economic investigators, including hospital and physician billing records and datasets from large managed care organizations (MCOs) or the Medicare program. Administrative datasets include information on resource quantities and can be used as a direct measure of resource utilization or abstracted to track utilization of specific services. As with medical records, certain administrative datasets may be available only for individual episodes of care (hospital datasets) and may not be available as linked records of inpatient and outpatient services. Also, Medicare and managed care administrative data may be available only for defined populations of patients and usually do not include information on noncovered services and specific service elements (for example, they may include little information on resources utilized during a hospitalization).

Patient Self-Report

Patient self-report enables data to be collected for all medical care received by patients, whether or not it is provided at the study site. Patient self-report data can be used in several ways in an economic study. First, it may be the only means of collecting certain data items (direct nonmedical costs, productivity costs, intangible costs). Second, patients are often the only people with complete knowledge of all of the services they have received. Thus, patient self-report can be used to assess direct medical costs received by patients. However, patient self-report data may not be reliable for several types of information, including data on resource utilization during acute-care hospitalizations or for detailed information on treatments received. In these cases, patient self-report information should not be used as the sole data collection strategy for economic studies. Often, patient self-report is combined with source-document validation. In this strategy, patients report the occurrence of a specific service at a facility, which then leads to follow-up data collection with that facility.

In addition to direct medical resource-use data, patient self-report may include assessment of the amount of caregiver time required for care at home and the number of days lost from work or other activities. Alternatively, work loss could also be abstracted directly from patients' work records, though it is unlikely that this would be feasible in a typical clinical trial unless the trial were restricted to patients with a single employer (for example, in a workplace cancer screening trial).

Several methods exist for collecting patient self-report data for resource use items. Patients may be interviewed, either during a protocol visit or over the telephone. Alternatively, specially designed questionnaires may be mailed to patients at regular intervals. Problems occur with patient self-report data for a number of reasons. Patient recall of events becomes problematic when the recall period is extended, if the patient is a heavy user of medical care services, and/or illness interferes with the patient's mental status. Thus, scheduled visits or telephone or mail contact must be sufficiently frequent to avoid recall problems. An alternative way to avoid recall problems is to ask the patient to complete a diary at home as care is received. The diary should then be brought to the study site at each visit, used by the patient as a reference during telephone follow-up, or used as a reference when completing a mail questionnaire. Diary problems can be minimized through reminder telephone calls to the patient at regular intervals between visits and by means of a letter sent before each study visit reminding the patient to complete the diary. For patients who are illiterate or mentally incompetent, resource data can be obtained from a proxy, such as a family member or close friend.

For an example of the use of combined patient interviews and medical billing records for obtaining direct nonmedical costs, see Bennett et al. *(40)*. Probably the most detailed and comprehensive survey instrumentation developed for the purpose of obtaining information on medical resource by way of patient interview is that developed for the National Medical Expenditure Survey *(41)*.

Assigning Unit Costs: Direct Medical Costs

Data on the costs of resources used by patients may be available from standard costing data, from sources of care for individual patients, or from administrative data sets. Most U.S. hospitals maintain billing systems that can be used to record the costs of resources consumed by patients during a hospitalization. However, hospital billing information will only include a record of the specific services supplied by the hospital and may not include the cost of physician services during the hospital period. Except in specific settings, hospital billing information does not include information on the care received by patients on an outpatient basis. Thus, if a protocol follows patients over an extended period of time, data collection mechanisms should be

available to collect both inpatient and outpatient resource use as well as the cost of these resources.

Administrative databases, such as the claims payment records of an MCO or large health plan, are becoming inereasingly important sources of resource-use data in clinical protocols. However, given the fragmented nature of the U.S. health insurance system, the use of administrative datasets is not a feasible means of tracking resource use unless the study is designed around specific patient populations, such as patients more than 65 years of age (Medicare) or patients enrolled in a specific health plan.

Hospital Costs

The most basic description of costs for hospital care is the cost of an entire hospital period. For example, a study in which the measure of resource utilization is hospital admission will assign an overall cost to each admission. The most common means of developing an estimate of aggregate cost is by first classifying the admission by diagnosis-related group (DRG), which combines the primary diagnosis and procedure with information about comorbidities. Medicare or other insurance payment information can then be used to assign a proxy for cost by the specific DRG. Another simple costing method tracks length of stay and assigns a cost per day, or a "per diem."

Generally, more specific information is collected on hospital resource utilization. Hospital financial information is available in the form of hospital charges. Because hospital charges are thought to overstate the actual costs of service, separate analysis of these data must be undertaken *(42)*.

Hospitals in the United States must report their overall costs and charges to the Health Care Financing Administration (HCFA) on an annual basis. This Medicare "cost report" has been used by investigators to develop a "cost-to-eharge ratio" that is used to convert hospital charges to hospital costs for the purpose of economic analysis. These cost-to-charge ratios can be developed either at an aggregate level for the institution or on a more specific level based on the hospital departments providing the relevant services. The relationship between costs and charges at hospitals actually varies considerably between hospital departments. When available, departmental cost-to-charge ratios may be a better proxy for costs of specific services than total cost-to-charge ratios. One severe limitation to this more detailed approach to assessment of hospital costs is that each hospital may have its own system of assigning costs to uniform billing (UB-92) categories. Thus, fresh frozen plasma may be included in one institution's blood bank cost center report but in the operating cost center in a different institution. These differences make aggregation of departmental cost-to-charge ratios a labor-intensive task. A study developed by the Cancer and Leukemia Group B (CALGB) is currently attempting to evaluate the capabilities of administrative datasets within CALGB member institutions. The group is expected to develop recommendations for strategies of requesting cost information from CALGB member institutions for economic evaluations within group studies.

Given the increasing finaneial pressures resulting from changes in the health care system, many hospitals have developed their own cost-accounting systems to provide more detailed cost information for management decisions. However, not every institution has such an accounting system in place. Where these datasets exist, they can be used to assess the costs of services and may offer a more accurate reflection of the true costs of services than the cost-to-charge ratio *(43)*.

It is difficult to develop costs for specific services from detailed hospital bills. Hospitals often keep track of every service received by an individual patient on a disaggregated basis. Thus, for a specific procedure or treatment—for example, an intensive care unit day or an hour of operating room time—the hospital may bill for each of the hundreds of components of that service separately. In fact, it may be impossible to develop an overall cost for specific services for an institution. One approach to resolving this issue is to develop a regression-based model that will use the resource counts in the economic protocol as predictor variables in decomposing the overall hospital bill. This technique could result in the development of bundled costs for the specific clinical services collected within a CRF.

More recently, hospitals have revised their billing procedures due to changes in market forces. For example, hospitals may agree to a contract price for procedures that are part of the research protocol, resulting in a gap between the actual bill and the contract price. Furthermore, some hospital accounting systems simply record the contract price. In these institutions, the process of tracking costs is predominantly being used for internal cost accounting but not for generating bills to external organizations. Some hospitals continue to generate bills but apply a contractual allowance to represent the difference between the charge and the contract price. Other hospitals have a separate accounting system for procedures priced through research contracts. If the data collection strategy for a proposed protocol is based solely on the use of administrative datasets, investigators must ensure that facilities are continuing to track resource utilization by specific patients in either the billing or cost-accounting systems at each center.

In a multi-institutional study, the costing strategy may be complicated by the inability to collect cost data at some study sites (most often because of lack of sufficient investigator resources to collect these data). In these cases, cost data may be obtained from a subset of study institutions. Statistical techniques to reflect this data collection strategy are discussed in Part V.

Physician Costs

Physicians assign CPT codes when billing for their services. In a manner analogous to hospitals, charges for specific physician services may not be directly related to the costs of providing those services. However, there is no physician cost-to-charge ratio. In 1992, HCFA developed the Resource-Based Relative Value Scale (RBRVS) as a measure of the resource intensity of the specific physician services for each CPT code. The Medicare system currently implements this resource intensity measure in physician payment. RVUs may be used to calculate standard costs for physician services *(44)*. It should be noted that there are facility charges associated with many outpatient medical procedures that are not included in the RBRVS fee schedule. These facility costs may need to be developed from other data sources. This is especially true for outpatient surgical services.

Nursing Services

Costs for nursing services are usually developed on an hourly basis using an average wage rate for the type of nurse (registered nurse, nurse practitioner, etc.).

Diagnostic Tests

Costs for diagnostic tests and procedures can include the costs of physician services as well as the costs of conducting the tests. The costs of physician services for diagnostic tests are reported as described above. The costs of conducting laboratory tests are available using a set of standard laboratory workload units (Lab RVUs). The costs of diagnostic tests include the facility fees for the service. The facility costs for radiology services may need to be developed from other data sources.

Pharmaceutical Services

Pharmaceutical charges also vary depending on the pharmacy. One way of standardizing pharmaceutical costs is to use the average wholesale price available for specific pharmaceutical products *(45),* with the addition of a standard dispensing fee.

Subacute-Care Facilities

Costs for subacute-care facilities may be developed based on published price lists using per diem rates. For studies that focus on the use of subacute-care facilities, a more detailed costing approach may need to be developed for the study. This would be the case in a study of alternative hospice treatment programs.

Capital Costs

Investigators should also consider that some treatments require the purchase of capital equipment; for example, eaneer centers may already own much of the equipment required for new treatments, but community hospitals may have to purchase equipment to implement certain services. Capital equipment purchases become fixed costs that are independent of the standard process of costing for an individual patient. In other words, there may be costs associated with a program of treatment that are shared by many patients, and this cost would not appear on patient billing data (this may especially be the case in phase III trials of experimental, unreimbursed technologies for which there is no market price for the service). Derivation of these costs requires specific costing efforts to be conducted in conjunction with the clinical trial.

Protocol Costs

The costs of data collection in a research study should not be included in the economic assessment of the therapy. In addition, the costs of research-related archived materials not used for clinical care should be excluded from the assessment.

Assigning Unit Costs: Direct Nonmedical Costs

Often, these costs are measured from the patient's perspective and include the costs of direct nonmedical services purchased by patients. While these data may be captured as both resources and costs, direct nonmedical costs are more often captured solely as costs of resources. This is usually assessed by asking patients to record their out-of-pocket expenses for these services.

Aggregating Price and Quantity

There are four general approaches to data analysis that investigators may consider in developing the economic component of clinical trials. These include resource use collected for all trial patients and costs collected for all trial patients; resource use collected for all trial patients and costs collected in a subset of study patients; resource use collected for all trial patients and costs developed from administrative data sets; and prospective collection of cost data only (resource utilization would then be derived from the cost data). In economic evaluation of cancer clinical trials, the second and third strategies are the most common.

Some economic evaluation strategies actually omit measurement of resource consumption and concentrate instead on economic measures of resource use, such as hospital bills. These studies are undertaken when collecting resource consumption measures may not be feasible or may be prohibitively expensive, forcing the investigator to choose an alternative "billing" approach.

Generic Economic Evaluation Strategies

Resource categories in clinical protocols must be specifically enumerated for both prospective and retrospective studies. The level of precision in the costing exercise is dependent upon the type of study being implemented.

Example. In a trial of low-molecular–weight heparin (outpatient) versus intravenous heparin (inpatient), no difference is expected in treatment outcome; however, a large difference is expected in the utilization of inpatient hospital resources. A list of resource categories to be collected in the clinical protocol might include hospital days by unit type, as well as physician visits.

Example. In a comparison of hospital settings in the administration of a 5HT3 odansetron bolus push versus intravenous administration, it is necessary to microcost the administration of the drugs. Such an approach represents a higher level of precision in the data than is required for the previous example. This study requires a microcosting exercise to track nursing time (and intensity) in providing antiemetic services to patients, with an effort to develop specific prices for these services.

Example. In a trial of adjuvant therapy for breast cancer assessing four versus eight cycles of chemotherapy, there is similar outpatient use of drug administration, pharmacy charges, diagnostic testing, and professional fees. The difference may lie in the hospitalization utilization rate for febrile neutropenia and in outpatient care for nausea, mucositis, dehydration, etc. Thus, investigators must evaluate outpatient resource utilization and occurrence of hospitalizations.

Inpatient Studies

Inpatient studies are clinical trials that either occur within the hospital setting or enroll a high proportion of patients that are expected to receive hospital treatment during the course of the clinical trial. Examples include trials of new cytokine therapies as supportive care for bone-marrow transplantation, and studies of new surgical techniques for solid organ tumors, studies of intensive chemotherapy regimens administered on an inpatient basis but potentially having different effects on bone-marrow

suppression. Data collection would focus on the intensity of treatment during the hospitalization, including tracking resource utilization by unit type within the hospital, physician visits, surgical procedures, diagnostic tests and procedures, and high-cost medications.

Outpatient Studies

Outpatient studies are clinical trials that either occur outside the hospital setting or enroll a high proportion of patients who are not expected to receive hospital treatment during the course of the clinical trial. Data collection would focus on the utilization of services on an outpatient basis, including physician services, nursing services, surgical procedures, diagnostic tests and procedures, and use of high-cost medications (including chemotherapy). Information about acute-care hospitalization would be recorded. However, unless there were an explicit hypothesis about the intensity of the hospital services by treatment arm, data collection on use of resources within the hospital setting could be minimized. For example, data could be collected on length of stay by unit type but not on diagnostic tests or procedures or medication use.

Subacute-Care Studies

Studies of patients within subacute-care facilities (i.e., hospice care, nursing home care, or rehabilitation care) require information about the use of the subacute-care facility but often do not require information about other outpatient treatments. Information about acute-care hospitalization would be recorded. However, unless there were an explicit hypothesis about the intensity of the hospital services with treatment, data on use of resources within the hospital setting could be minimized.

Conclusion

In summary, this section addresses some of the major issues in designing and implementing an economic evaluation within a clinical trial. Issues highlighted in this section include the potential direct medical and direct nonmedical resources that can be captured in a clinical trial, the assessment of productivity and intangible costs, the sources of data for capturing resource utilization information, the methodology for assigning unit costs to the resources collected in the protocol, and methods of refining data collection instruments for different types of clinical trials. There is no "boiler-plate" economic evaluation. Clinical and economic investigators must review the requirements for the different data elements within their protocol. They must also assess alternative strategies for collecting cost information for the resources collected within the economic protocol.

Part IV: Designing an Effective Data Collection Strategy

Designing Data Collection Forms

When integrating economics into an existing clinical trial design, investigators should attempt to collect only those items that are necessary for the economic analysis. Existing forms, such as clinical trial report forms and flow sheets, often contain much of the required information (*see* Appendix I). It may be necessary to add only a few variables to collect sufficient information to conduct the study. Variables should include any subset of those listed below, depending on the type of trial and the study analysis plan.

When collecting data centered around actual resource use, the CRFs used to collect the data must be designed and pilot tested prior to data collection. They should allow the investigators to record core areas of resource use required for the study. Specific forms are developed to track major resource use by study patients. These forms often are divided into hospital resource use, outpatient resource use, other institutional resource use, caregiver support, and covariates.

Hospital Resource Data

The following are important data elements to consider for collection when patients are hospitalized during the clinical trial:

- facility name and location;
- admission and discharge dates;
- reason for admission (e.g., protocol treatment, elective admission, and urgent admission);
- hospital room type (e.g., ward, step-down, intensive care unit);
- surgical procedure (type as well as duration of procedure);
- transfusions of blood or blood products (date, type, number);
- major laboratory tests (date, type, number);
- major radiologic tests (date, type, number);
- medications (date, type, route, dosage);
- chemotherapy (dates, types, routes, dosage);
- radiotherapy (dates, fractionation, site, dosage); and
- physician visits (date, type, duration in minutes).

Outpatient Resource Data

The following are important data elements to consider for collection when patients are treated in an outpatient setting:

- date, type, and duration of visit;
- type of facility;
- reason for visit (routine, elective follow-up, emergent follow-up);
- surgical procedure (type and duration of procedure);
- major laboratory tests (date, type, number);
- major radiologic tests (date, type, number);
- medications (date, type, route, dosage);
- chemotherapy (date, type, route, dosage);
- radiotherapy (date, fractionation, site, dosage);
- physician visits (date, type, duration in minutes); and
- home care (dates/frequency of visits, skill requirements).

Subacute-Care Institutional Resource Use

The following are important data elements to consider for collection when patients are treated in subacute-care institutions:

- nursing home (date of admission, date of discharge, level of care); and
- hospice care (date of admission, date of discharge, level of care).

Caregiver Support

The following are important data elements to consider for collection when assessing caregiver burden:

- caregiver burden (number of hours of paid and unpaid caregiver support);
- days of usual activity missed (work, school, home activities);
- out-of-pocket costs;
- transportation costs; and
- caregiver quality of life.

Covariates

The following are important covariates to be collected within the economic trial:

- comorbidities;
- patient's demographic characteristics;
- geographic proximity of patient to site of care;
- insurance coverage status;
- caregiver status;
- prior hospitalizations (within the past year);
- employment status;
- disease severity;
- prior physician visits (within the past year); and
- cancer history (prior chemotherapy, prior radiation therapy, prior surgery, date of diagnosis).

Collecting Data

Beginning Data Collection

There are essentially three approaches to collecting trial data on a prospective basis—1) collection at a fixed time period; 2) collection at the beginning of a designated medical event; and 3) collection after a treatment cycle or hospitalization. The strategy for data collection should be planned well in advance and will depend on the needs of the specific trial (for example, whether outpatient treatment is being compared with inpatient treatment).

Collection at fixed time periods. Most economic studies query patients about resource utilization at fixed time periods—for example, on a monthly or quarterly basis. This method assures that economic data are not omitted in the study protocol and that follow-up is comparable across the study areas. This data collection strategy is also the best method to assess quality-of-life data for study patients.

Starting from an event. Clinical trials that measure resource utilization from the time of a predesignated event often use toxicity as a marker. For example, a data collection form might instruct, "In the case of toxicity, please indicate (on another form) what resources were used to treat the toxicity." The event-triggered method is limited in that a significant amount of resource utilization may not be linked to toxicity.

Starting after a treatment cycle or hospitalization. Data can also be collected at the end of a treatment cycle, such as a cycle of chemotherapy, or after a computed tomography (CT) scan or other procedure or test. This method provides reasonable certainty—if interviews are conducted at least quarterly—that all instances of resource utilization and the reasons for each will be captured. More frequent interviews may be required for sicker patients, though the data collection strategy should be standardized to keep costs low.

Why Collect Nonstudy-Site Data?

Although nonstudy-site data are difficult to collect, the data do reflect resource utilization at sites outside the treatment center. The costs of these resources ean be substantial when significant portions of care are provided in these settings. For example, bone-marrow transplant patients may receive much of their follow-up care at nonstudy sites.

Obtaining Informed Consent

Primary economic data collection that requires access to medical records or other confidential sources must have the clearance of the study subject. Trials that follow patients to nonstudy centers must develop more elaborate consent procedures to ensure that all patient records are included in the consent. In general, a consent form should be designed to encompass as many potential sources of information as possible, particularly if the investigators anticipate having to collect information from outside the study center. The consent form should allow the study investigators access to the patients' financial records at whatever sites the patients receive care.

Medical resource-use data collection may involve abstraction by trial personnel or data from the patient's medical records or medical bills. To obtain a patient's medical bill in the United States, the principal investigator must include a release-of-information statement in the patient consent form. This can be a requirement for enrollment in the study or optional depending on the design of the study. Patient trial numbers should be substituted for patient names to maintain anonymity for all confidential data sources for analysis, especially on financial data. Strict confidentiality procedures should be applied to all economic data.

Maintaining Communication Between Research Teams

Most large, multisite clinical trials are monitored at the sites by a team of clinical research associates. These individuals play an important role in maintaining high data quality for both clinical and economic data. Often, they are more experienced at reviewing clinical data than economic data; therefore, it is usually important to conduct training sessions at the beginning of the trial in both the general requirements of the economic study and the specific review that should be conducted. Those items for which source-document review can and should be done must be specified in advance; often the clinical research associate will be a good judge of what is feasible. This training must be followed by regular communication between the economic investigators and the clinical research associate during the data collection period to correct problems as they occur.

Conclusion

In this section, we review in more detail the specific data elements that often are considered for inclusion in an economic evaluation. We further discuss procedural issues, such as the need for informed consent and for maintaining effective communication within the research team. Economic evaluation requires a comprehensive approach to the development of data collection instruments and a data collection strategy. This section presents a general framework for this assessment.

Part V: Statistical Issues

In this section, we aim to describe the unique characteristics of economic data as they are encountered in controlled clinical trials; we intend, furthermore, to identify a number of the statistical complications associated with such data. We suggest a variety of statistical methods that might be used to address these complications. This chapter is not meant to be taken as a cookbook that offers a standard menu of tools that can be uniformly applied to any economic data collected in the course of controlled clinical trials. Rather, it is meant to alert trial planners and analysts to the statistical concerns that may need to be addressed in the design, implementation, and analysis of trials oriented toward economic issues and to provide some guidance to the rich and growing statistical literature that has been engendered by the increasing practice of collecting economic data alongside clinical trials.

The Importance of a Statistical Analysis Plan

The same care and rigor must be applied to the statistical analysis of economic data as is routinely applied to clinical endpoints in clinical trials. Detailed analysis plans are as important for economic evaluations as they are for clinical evaluations. The primary economic outcome should also be specified as closely as possible, as should covariates that are intended to be included in analysis for descriptive or predictive purposes. Included in the study protocol should be a specification of the statistical strategies that will be used. In deciding on appropriate statistical methods to use, such features of the data, such as its distributional characteristics, the degree and type of censoring, and other types of missing data, should be considered. A properly conducted sample-size analysis should be conducted for economic data. Also, the analyst should consider whether economic data from a clinical trial alone are sufficient for addressing the economic question of interest or whether such data must be supplemented with data derived from other sources or from modeling or extrapolation techniques.

Many of the statistical issues that apply to clinical endpoints are also applicable to economic data. However, economic data are in some ways different and more complicated than clinical data, and additional issues may arise in regard to the analysis of economic data. These characteristics of economic data should be considered before deciding on a strategy for statistical analysis.

Important Properties of Economic Data

Anticipated Uses of Economic Data

Clinical endpoints often are reported as ratio terms; for example, treatment A resulted in a 30% mortality reduction compared with treatment B. It is also possible to report costs such that average medical costs for treatment A were 20% greater than for treatment B. This approach may be appropriate when the only aim of the study is to test the null hypothesis of equal costs between two treatment arms.

It is often anticipated, however, that the economic results of clinical studies will be used in economic analyses in which the absolute difference of cost is of interest. So, in a cost-effectiveness or cost-utility analysis, average medical costs for treatment A might be $10 000 greater than cost for treatment B. The same point applies to the clinical outcome in cost-effectiveness and cost-utility analyses. That is, an estimate of absolute, rather than relative, benefit is required for the denominator of the cost-effectiveness or cost-utility ratio. For example, patients receiving treatment A survived 2 years longer than patients receiving treatment B. The anticipated use, in economic analysis, of a ratio estimate, such as the cost-effectiveness ratio—consisting of the ratio of the absolute cost difference divided by the absolute treatment effect difference—may have implications for the statistical design of a trial in which economic considerations are prominent.

Economic analysis may be focused on estimating the economic impact of a treatment for a patient group or treatment setting that is not typical of the trial setting. In such cases, covariate analysis becomes essential. Because different covariates may be more important for economic effects than for clinical effects, it may not be possible to control for these covariates through such strategies as block randomization. Covariates that are of interest in economic analysis and may or may not be controlled for in clinical trials include those reviewed earlier, including sociodemographic variables (sex, age, race, immigration status, income, education, occupation, employment status, census tract of residence, and marital status), comorbidity status, functional status, treatment setting prior to randomization, treatment setting before initial trial care, cost prior to randomization, availability of caretaker at home, availability of social support networks, type of health insurance, and structure of health care provider organization.

The Complexity of Economic Data

While a variety of clinical measures may be recorded in trials and used in complex ways in covariate analysis, primary outcome measures in clinical trials often are measurable by a single index, such as death from or recurrence of cancer, etc. By comparison, economic data—not unlike quality-of-life data—tend to be complex.

First, economic data usually are not measurable as a single event but as an accumulation of resource use over time. Second, economic data may consist of a variety of components that may, in turn, serve as substitutes or complements to each other (i.e., positively or negatively correlated). For example, costs may consist of the components of hospital-based costs, outpatient costs, and family costs. Costs may be equivalent for patients A and B, but patient B's outpatient or family costs may have been substituted for hospital costs. Third, the longitudinal nature of costs present specific analytic issues. For example, if certain components of cost have a systematic tendency to be measured with less complete data in treatment arm B, and this is not recognized in the analysis, an incorrect inference could be drawn that costs are lower for B.

Within these broad components of cost, cost is measured as a collection of finer components. For example, hospital costs may consist of the summation of many procedure or service components, such as operating room use, intensive care days, normal hospital days. The case is similar for outpatient care. Because cost components may be substitutes or complements, it is important that collection of all relevant cost data be as comprehensive as possible. In addition, it follows from this property of costs that tests of statistical significance should not be performed on individual components of cost, nor should individual cost

components be excluded from the summation of costs because of statistical insignificance *(46)*.

The second way in which economic data are complex is that cost is the sum of products of resource use (quantity) and unit cost (price). Cost may not always be separated into its component quantity and price vectors, but often it will be, because there is great uncertainty about the validity of the price vectors that are embedded in the usual accounting systems available from hospitals and physician office records. When the quantity and price vectors that compose cost are obtained separately, the question of how to treat the separate and joint variability of these quantity and price vectors arises.

Because economic data tend to be complex, the statistical analyst is more likely to have to deal with the problem of missing data. This is due both to the variety and complexity of data elements necessary to measure economic variables and to the fact that economic data elements may be drawn from administrative data systems that are not designed primarily for research purposes. Also, to the extent that some data elements are derived from questionnaires administered to patients or caregivers, missing data will be a problem.

Distributional Characteristics of Economic Data

Cost data may have a relatively skewed distribution. An idea of what the distributional characteristics of cost data from cancer clinical trials might look like can be inferred from observational cost data from the SEER[1]–Medicare database *(47)*. This database links records of incident cancer cases from NCI's SEER system to claims data from the Medicare system maintained by HCFA.

Tables 5 and 6 show the distributional characteristics of actual payments made by Medicare for treatment of colorectal cancer by stage of diagnosis, covering all costs from the date of diagnosis to 6 months following those dates (only patients surviving at least 6 months from the date of diagnosis are included in the analysis). Table 5 shows data using charges submitted to Medicare as a measure of cost, and Table 6 shows data using actual reimbursements paid by Medicare as a measure of cost. In all cases, these data do not fit the assumptions of normality, and for all cases except one—reimbursements for distant-stage disease—the data tend to be more highly skewed and dispersed than data from an exponential distribution.

Actual economic data encountered in clinical trials may be less skewed than the data displayed in Tables 5 and 6 because the population and treatments studied in the context of clinical trials may be more homogeneous. On the other hand, economie data encountered in clinical trials may be more skewed because, unlike the SEER–Medicare data, it may be derived from a wide variety of accounting systems. Also, the inclusion of additional components of cost, such as patient out-of-pocket or time costs, may introduce additional components of variance and skewedness. Dudley et al. *(48)* present another example of highly skewed cost data from a clinical trial of coronary artery bypass graft surgery.

Table 5. Distributional characteristics of SEER/Medicare Data: submitted charges for colorectal cancer—cumulative charges to 6 months from diagnosis*

Stage	Mean	Median	Maximum	Skewedness	Kurtosis	Standard deviation
In situ	$13 009	$ 6072	$300 670	4.68	40	$18 943
Local	$23 821	$18 765	$665 256	4.47	50	$23 237
Regional	$29 336	$22 961	$730 504	4.60	54	$25 909
Distant	$33 207	$26 731	$525 191	4.25	43	$26 731

*Source: SEER–Medicare database, Applied Research Branch, National Cancer Institute.

Table 6. Distributional characteristics of SEER/Medicare Data: Medicare reimbursements for colorectal cancer—Cumulative reimbursements to 6 months from diagnosis*

Stage	Mean	Median	Maximum	Skewedness	Kurtosis	Standard deviation
In situ	$ 8529	$ 4244	$135 991	2.74	18	$10 141
Local	$16 049	$16 082	$384 138	3.12	61	$11 030
Regional	$19 707	$18 605	$295 032	2.82	33	$11 297
Distant	$21 425	$20 461	$105 975	1.02	4	$12 221

*Source: SEER–Medicare database, Applied Research Branch, National Cancer Institute.

Statistical Methods for Analyzing Economic Data

Descriptive Methods

Mean costs are perhaps the most important univariable summaries, and in the absence of censoring problems, they can be simple arithmetic means. Means are important because their multiplication by population size is an estimate of total system costs, and the comparison of total system costs conditional on comparison treatments can have economic meaning for policy evaluation. However, when cost data are heavily skewed, mean costs may not be descriptive of typical patients, so median costs and other quantiles should always supplement the presentation.

The standard deviation of eosts may not be very informative, as its interpretation depends on normality of costs (one frequently sees illogical published figures such as "the mean cost was $10 000 plus or minus $12 000"). An alternative set of summary statistics would be the mean, median, and quantiles such as 0.05, 0.1, 0.25, 0.75, 0.9, and 0.95. Graphical presentation of the empirical cumulative distribution function of costs is even more informative.

Univariable Methods

Analysis of economic data without informative censoring. In some trials, the economic data of interest derives from a single event or a series of events unrelated to the subsequent flow of costs or survival. Also, cumulative costs over a fixed duration from the date of randomization are of interest, and observations on costs up to this date are complete. In cases such as these, there will be no informative censoring. A variety of statistical methods applicable to cost data are available in this case.

Testing the entire distribution. As stated above, the difference in mean costs will be of most interest. However, it is useful to conceptualize cost comparisons by comparing the entire cost distribution between treatment groups. The most general method for comparing distributions is the Kolmogorov–Smirnov test, which is based on the maximum absolute difference between

two cumulative distribution function estimates. If the sample is large enough to make relative efficiency a nonissue, the robust Kolmogorov–Smirnov test would not have many competitors. However, for moderate to small studies, this general nonparametric test does not have as much power as tests directed at detecting differences in specific parameters of the distribution, such as the mean or median.

Tests for difference on means, medians, and other parameters. For directed tests, one can use either parametric or nonparametric tests. The latter may be preferred because 1) if normality holds, nonparametric tests such as the Wilcoxon test still have 0.96 efficiency with respect to the optimal test (Student'*t* test); and 2) if normality does not hold (the likely condition for cost data), nonparametric tests can be much more efficient than the Student'*t* test. For symmetric distributions, the Wilcoxon two-sample test addresses whether the mean or median cost differs between groups, because under symmetry the mean equals the median. For the general case in which the distribution may be asymmetric, the Wilcoxon test is still a valid test for differences in central tendency. As such, the Wilcoxon test is most sensitive in test of the pseudomedian, defined as the median of midpoints of all possible pairs of costs. More important, however, the Wilcoxon two-sample test is a valid test of the null hypothesis that the probability that costs in group A exceed costs in group B, for randomly chosen pairs of patients, regardless of the distribution.

The Wilcoxon statistic is a simple translation of an estimate of the probability that a random A patient has incurred costs exceeding those of a random B patient. Thus, the Wilcoxon test has many advantages for univariable resource-use comparisons. Other nonparametric tests, such as the log-rank test, can also be considered. The log-rank test has the advantage of being a special case of the Cox semiparametric regression model, which may come into play when covariable adjustment is needed.

When the log of costs is thought to be normally distributed, log-transformed comparisons testing for differences in the median cost (equivalent to testing for differences in mean costs if the group variances are equal) have often been used. Zhou et al. *(49,50)* have developed a method, the Z-score test, to test for differences in means when group variances are not equal by recognizing that the log of the mean of the untransformed data is equal to the mean of the log-transformed data plus one half the variance of the transformed data. The Z-score test gives type I errors closest to nominal levels when variances differ between the two distributions being tested.

The validity of parametric tests, such as the Z-score, depend on the correctness of the assumption about the distributional characteristics of the data, which may be difficult to establish with relatively small samples. One way to obviate this difficulty is to use tests whose validity does not depend on the distribution of the data, such as permutation *t* tests *(51)* or bootstrap techniques *(52)*. These approaches are preferred over the Z score when it is uncertain that the distribution of log-transformed cost data approximates a normal distribution. The disadvantage of permutation tests and the bootstrap technique is that these methods generally are not available in the form of "canned" software packages and, therefore, require a moderate amount of custom programming by the analyst. The bootstrap technique is further described later in this section.

Covariable Adjustment and Multivariable Modeling Methods

In general, two reasons for using covariable adjustment in the context of a randomized clinical trial are to gain power to shorten the confidence interval (CI) for the treatment effect and to estimate costs and health effects for important subgroups when trial participants are heterogeneous. Some authors have found that for binary endpoints or censored failure time outcomes, covariable adjustment may not result in any appreciable gain in precision or power *(53,54)*. However, in heterogeneous patient populations, estimating crude effect ratios (unadjusted odds or hazard ratios) can result in different estimates than adjusted estimates, and the crude estimates may not apply to any patient. Also, unadjusted models may lack fit. For example, the treatment may operate in a nonproportional hazards fashion without adjustment but satisfy the Cox model's proportional hazards assumption with adjustment *(55,56)*. When patients have a moderately wide prognostic spectrum, unadjusted tests (e.g., unstratified log-rank test) may have poor power in comparison with adjusted tests (e.g., from the Cox model) *(57)*.

In economic analysis, the outcomes of interest are the absolute cost difference and the absolute effects of treatments. Even when power and precision are not issues, modeling multiple patient descriptors or prognostic factors is necessary when one wishes to translate relative treatment benefit into absolute benefit. It is easy to demonstrate that, even when relative treatment effects are constant across all types of patients, absolute effects vary when the variation of prognostic factors is moderate across enrolled patients. The absolute survival advantage of a treatment is strongly dependent on initial survival in the absence of treatment as well as the relative survival advantage associated with the treatment.[2] Thus, covariable modeling is necessary in estimations of absolute effects of treatment on event probabilities or survival time that comprise the denominators of cost-effectiveness ratios, and covariables may also be needed to increase power.

Similarly, for cost endpoints, one expects that covariable adjustment will almost always increase power, and it will also be useful for estimating cost distributions (or point estimates and confidence limits) for different types of patients.

Regression models. Ordinary multiple regression currently is the most commonly employed regression model for analysis of costs. These models can be used to derive unbiased estimates, but when the distribution of cost data is highly skewed, the estimates may have high variance, and confidence limits for mean costs may not have the desired coverage probabilities. As discussed briefly above, one of the traditional methods for regression analysis of costs is to model the log of costs. For example, Rutten-van Moken et al. *(58)* use a log transformation of costs in their analysis of economic data after first applying a linear transformation to deal with the problem of zero costs. They perform regression analysis on the log-transformed costs and then use the Duan "smearing" estimator *(59)* to retransform the regression-adjusted cost estimates back to the linear domain. This method may not be entirely satisfactory because the residuals may still not be normally distributed, and the assumed multiplicative relationship between patient variables and cost (on the original scale) may not be sensible. For example, analyses of data from a clinical trial on the cost of coronary artery bypass

graft surgery found that one would have to take sixfold logarithms of charges to achieve normality *(48)*.

An alternative to using log transformation models is to use robust regression techniques. The Cox semiparametric regression model *(60,61)* has demonstrated several advantages in analyses of resource-utilization data *(48,62,63)* characterized by a skewed distribution. Regression coefficient estimates for the Cox model are identical whether one analyzes costs on the original scale or takes any monotonic transformation of costs (such as the logarithm). Antilogs of regression coefficients are hazard ratios, but these can be converted to cost ratios *(64)*. Once regression coefficients are estimated, the entire cost distribution, or the mean, median, or any other quantile of cost, can easily be estimated for individual types of patients. By assigning treatment as one of the predictors, one can readily estimate differences in mean or median costs using the fitted model. As the Cox model is distribution-free for any single type of patient, there is no *a priori* assumption about how the mean and median are related mathematically. Model fit can be checked using standard methods, i.e., smooth-scaled Schoenfeld residual plots for checking the proportional hazards assumption, and smoothed martingale residual plots or regression spline fits for determining linearity of predictors *(64)*. Note that the Cox model in no way assumes that any censoring is present, so it can be used to analyze continuous response variables of many types.[3]

It should be recognized that while the standard Cox model is often used to analyze clinical outcome data in the presence of censoring, this approach to censoring may result in bias in the cost realm (although the bias resulting from ignoring censoring may be even greater). This issue is discussed in more detail below.

Analysis in the Presence of Informative Censoring

Cumulative costs, which are the integral of the flow of costs over time, often are the economic data of primary interest in clinical trials. This flow may not be constant over time, and in many eases, the flow will increase beginning sometime prior to death. Observations of this cumulative flow from a trial may include patients with various times accrued since the time of randomization and may include some patients who have died, others who are close to death, some who will not die for many years, and some who are lost to eeonomic follow-up at various times from randomization.

Survival Techniques

Standard techniques of survival analysis have been recommended for analyzing cost data in clinical trials in the presence of censoring. For example, Fenn et al. *(62)* have recently advocated the use of the Kaplan–Meier estimator for comparing costs in clinical trials. This approach has obvious appeal because it is parallel to the usual method of analyzing clinical survival in cancer trials.

However, the scale of cumulative cost is not strictly equivalent to the scale for the clinical endpoint of time. Time has a constant rate of flow while cost does not. In addition, changes in the flow of costs are not random; rather, increases in the flow of costs tend to be indicative of death. Thus, when cost data are censored prior to death, this censoring is informative with regard to costs and survival. Even when no deaths are present in the data, the different scales for cost and censoring can cause informative censoring.

A number of variations on the Kaplan–Meier estimator have been proposed to take this complication into account. Etzioni et al. *(65)* have proposed the Kaplan–Meier sample average (KMSA) estimator. For each interval, the KMSA estimates the sample average cost for a given time interval as the average cost of all individuals who were alive and observed at the beginning of that period. Total cumulative cost is then given as the weighted sum of all interval costs where the weights are the Kaplan–Meier estimates for the probability of survival up to that interval (since the true probabilities are not known because of censoring). This estimator works well when 1) cost data are available at frequent intervals; and 2) the censoring mechanism is independent of survival and cost in the sense that individuals are not censored because they appear to be at especially high or low risk of death. In addition, individuals are not censored because of accumulation of especially high or low costs relative to uncensored cases.

A variation on the KMSA estimator is the Pathwise estimator *(66)*, which requires only data on cumulative costs for individuals who have died (or reached some other "terminal" follow-up point, such that the flow of costs at that point are not correlated with the probability of death). The Pathwise estimator of average cumulative cost is the weighted sum of the average cumulative costs (of individuals who died) in each interval, where the weights are the Kaplan–Meier estimates for the probability of survival up to that interval. The Pathwise estimator does not depend on the availability of cost data for short intervals (data based on annual intervals can be used), and it yields an estimate of the distribution of cumulative costs as well as the mean. The efficiency of the Pathwise estimator is limited by the proportion of cases in which death is observed. A variant of the KMSA/Pathwise approach has been used to estimate cumulative cancer treatment costs using Medicare data linked to the SEER tumor registry system by treatment phase *(67)*. Efforts are currently underway to extend the KMSA/Pathwise approach to Cox regression *(66)*. Etzioni et al. *(68)* use SEER–Medicare data combined with simulated censoring to demonstrate the bias in estimating costs when using standard Kaplan–Meier and Cox regression techniques and the improvement achieved by using the KMSA and pathwise methods. Etzioni et al. *(65)* has discussed the bias that may be introduced in cost extrapolations using the KMSA approach when costs are discounted.

Parametric survival models have also been advocated as a method for extrapolating costs beyond the endpoint of trial observation *(69,70)*. Fenn et al. *(62)* describes this approach using standard survival techniques. The results of parametric extrapolation may be very sensitive to the specification of the survival function and, within trial data, may be inadequate to allow strong discrimination between specifications. In this case, the variance introduced into cost predictions by model specification uncertainty should be recognized *(71)*.

Missing Data

Several types of missing-data problems are likely to occur in relation to economic data collected in clinical trials. The types of missing data discussed in this section are different from the types of missing data that can be dealt with using the survival methods

reviewed above. This section discusses data that are missing due to nonresponse or a failure to collect or record specific data items due to technical or logistic problems—not as a result of censoring inherent to the study's time frame.

Item Nonresponse

Item nonresponse occurs when one or several of the elements of an outcome variable at a specific point in time are missing but when there is still a great deal of data available on the patient. For example, data on the number of physician visits for a patient might be missing for a specific week but be available for all other weeks; or unit cost data for a specific procedure may be missing from the accounting system of one study hospital but available at others. Item nonresponse can occur at intermediate points in time, in which observations exist before and after the missing item, as well as at the point of a final outcome. Because item nonresponse implies that there is sufficient information on the patient already available, it is reasonable to include the patient in the analysis despite the missing elements. It is acceptable to impute missing data when the analyst believes that the study will not be biased by imputing the missing values.

Unit Nonresponse

Unit nonresponse occurs when all or most of the information is missing on the unit of observation, such as all hospital days for a particular patient or all unit cost data for a particular study hospital. One type of unit nonresponse, in contrast to censoring caused by the time frame of the study, is the type of missing data not handled well by the survival methods discussed above (i.e., when the cost histories of patients who "drop out" are correlated to the reason for dropping out and are very different from patients who do not drop out).

Imputation Methods for Missing Data

Methods of imputation for missing unit data are reviewed by Little and Rubin *(72)* and Rubin *(73)*, and specific application to health care data is described by Rubin and Shenker *(74)*. Imputation strategies range from using the last observed value of the missing item or unit to using mean values of the missing item or unit to modeling the missing values using regression or simulation models that link the missing values to observed data. The appropriate imputation strategy depends on the analyst's judgment regarding the nature of the missing data. If the missing data are believed to be missing completely at random (MCAR) (i.e., missing data are independent of observed data or unobserved data if the latter were known), rather simple methods of imputation are appropriate. If data is not MCAR, but the probability of missing is related only to observed data, then various structural imputation models may be appropriate. Lavori et al. *(75)* apply Rubin's method of multiple imputation to unit data missing from a clinical trial due to patient nonadherence. They then compare the results to those obtained using the simpler last value approach and using only complete data records. Multiple imputation gives values of clinical outcomes that are between the extreme values of the other two methods and yields higher (more realistic) variances. Bootstrapping can also be used to compute variance estimates and CIs that take into account the increased variance resulting from having imputed missing data *(76)*. McGuigan et al. *(77)* compared the results of the two-stage Heckman sample selection model and the results of inverse sample weights based on logistic regression to results actually obtained when additional follow-up data were obtained for a cohort study on a school-based drug prevention program. The weighting strategy performed better than the sample selection model and better than analysis of data without the additional follow-up. Performance of the sample selection model was very sensitive to the particular specification of the model used.

When billing or cost information are not available for all patients, cost data can be collected on a subset of patients with resource utilization data being collected for the entire sample. Regression techniques can then be used to impute charges and predict bills. It should be remembered that the selected institutions from which bills are obtained are not necessarily representative of other institutions. For example, institutions with the best billing systems may also have more efficient delivery care systems. Or, the billing systems upon which unit costs are imputed may have higher unit costs, because they are less efficient and have less advanced billing systems. To prevent bias, resource utilization must be compared across institutions.

Although much time and energy are involved in collecting bills at every institution, it is recommended that an effort be made to obtain complete billing data for major resource utilization events when possible. The likelihood that the distribution of expenses for an institution with missing billing data are different than other institutions should be addressed. If a difference in resource use is found, costs from the hospitals that seem to be driving those differences in resource utilization should be obtained.

How should the collection of data for both resource consumption and pricing studies be handled for hospitals that may be in pricing consortia and will not disclose the prices of some resources or certain pharmaceuticals? A price may have to be assigned. For example, a military institution, which is a large accruer to a number of trials, will have a different price on almost every resource than a nonmilitary site. Access to technology and length of stay may radically alter pricing information among institutions. Therefore, one may only be able to assign a reasonable across-the-board price in the estimate. This must be explicitly stated in the analysis. Speaking directly to Internal Review Boards (IRBs) concerning the financial use of the data, especially in terms of confidentiality, may assist in disclosure of needed information.

Variance of Price and Quantity Vectors

In some studies, direct measures of monetary cost will be available. In others, costs will be constructed by multiplication of price (unit cost) vector and a quantity (resource utilization) vector. In the latter situation, data on resource utilization often will be obtained from observations on study subjects, while data on the price vector will be obtained from data sources not directly related to these observations. In this case, it will be difficult or impossible to determine the variance of the price vector and the covariance between the price and quantity vectors. In this case, it is permissible to treat the price vector as if it were a parameter of fixed values. But, it should be recognized that this will result in an optimistic estimate of the variance of cost. One way to account for this is to perform a sensitivity analysis in which the values of the price vector are varied over a "plausibility interval." Upper and lower bounds for the price values

might be determined from observational data, such as billing, reimbursement, and cost-accounting records of various health care providers or third-party payers.

Statistical Issues Related to Cost-Effectiveness Ratios

What Should Be in the Denominator?

The typical quantity to use in the denominator of a cost-effectiveness ratio is the increase in expected (mean) life length, considering death due to all causes. There is nothing inherently wrong with using the increase in median life length, and this quantity requires less extrapolation and in some cases does not require parametric survival modeling. Another choice is the mean restricted life length, the area under the estimated quality-adjusted survival curve from zero until the end of the follow-up period of interest *(78,79)*. This measure also eliminates the need for extensive extrapolation. Another choice for the denominator is the difference in t-year survival probabilities. Mark et al. *(80)* compared many of these measures in one setting. Similarly, the denominator for a cost-utility ratio is the life length adjusted for quality.

Variance of the Cost-Effectiveness Ratio

When economic and clinical outcome data from clinical trials are used to construct a cost-effectiveness (or cost-utility) ratio, the statistical properties of this kind of ratio estimator should be appreciated. The full variance/covariance structure of costs and clinical outcomes will usually not be known. A variety of techniques have been developed to provide estimates of the variance of the cost-effectiveness ratio in this situation. These are reviewed by Manning et al. *(81)* and Polsky et al. *(82,83)*. These techniques include univariate sensitivity analysis, multivariate sensitivity analysis, computation of confidence regions, and overall CIs using the delta method or simulation techniques. Naïve use of univariate analysis will most often result in estimates of CIs that are unrealistically narrow, and multivariate sensitivity analysis may result in CIs that are too wide if only extreme bounds are examined. Manning et al. *(81)* compare the results of univariate and multivariate sensitivity analysis, the delta method and Monte Carlo simulation using an empiric example of the cost-effectiveness of breast cancer screening, where cost-effectiveness is a function of nine epidemiological, teehnical, and economic parameters.

Application of the bootstrap method to obtain joint variance plots of the numerator and denominator of the cost-effectiveness ratio and to perform threshold analysis is described by Hlatky et al. *(84)*. The bootstrap can also be used to simultaneously take into account other sources of variation such as missing data and model selection *(85)*.

Interpretation of the confidence interval from a bootstrap assessment is based on the proportion of times a result falls into each of the four possible quadrants of the cost-effectiveness plane, as well as an assessment of how often the cost-effectiveness ratio is below a prespecified threshold (e.g., $50 000 per QALY gained).

Fig. 1 provides an example of the results of a bootstrap procedure that was used to estimate the confidence intervals for the comparison of costs and effects of a new therapy and usual practice. The *x* axis represents the difference in mortality between the two therapies. Points to the right of the line representing zero (i.e., no difference) indicate that the new therapy increased the probability of survival, while points to the left of the line indicate that the new therapy decreased the probability of survival. The *y* axis represents the difference in costs between the therapies. Points above the line representing zero indicate that the new therapy increased costs, while those to the right indicate that it decreased costs. The diagonal line represents the combinations of changes in costs and effects that bear a ratio of $50 000 per death averted, one example of a prespecified level of cost-effectiveness that might be proposed to justify new investments.

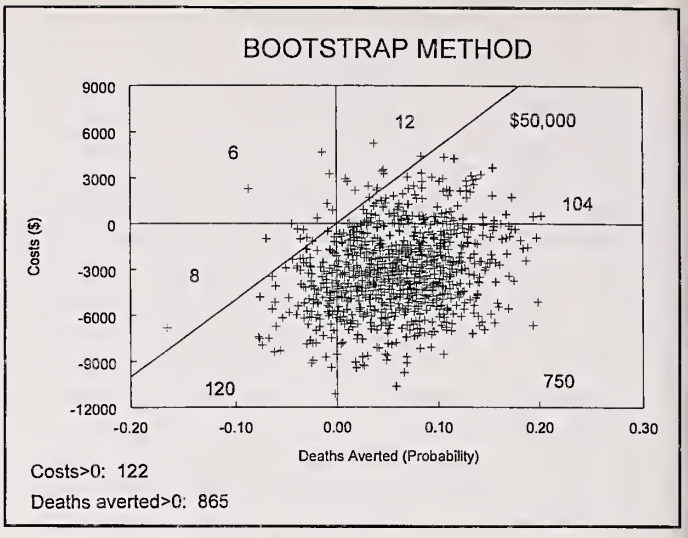


Fig. 1. Imputation of missing data.

In quadrant I of the Fig. 1, the 104 points below and to the right of the line represent cases in which the new therapy has a ratio below $50 000 (i.e., satisfy the acceptability criterion), while the 12 points above and to the left of the line represent cases in which the new therapy has ratios that fail to satisfy the criterion.

In quadrant III, the new therapy is less costly and less effective. The 120 points below and to the right of the $50 000 line indicate instances in which the use of the more expensive, more effective usual practice has ratios that exceed the $50 000 ratio (i.e., the new therapy would save enough money to potentially justify the trade-off in health), while the 8 points above the $50 000 line indicate instances in which the adoption of more expensive, more effective usual care would have ratios that satisfy the criterion.

In all, in 974 of the 1000 replications, the new therapy had a change in costs and effects that satisfied the $50 000 criterion (i.e., given a two-tailed test, we can conclude that there was just over a 5% chance that the new therapy satisfied the acceptability criterion). While there is evidence that the therapy satisfied the $50 000 acceptability criterion, the statistical tests for the individual comparisons of cost and mortality differences between the two therapies indicate that there were no differences between the therapies at the 95% CI (in 122 of 1000 cases, the costs were increased by the new therapy, while in 878 they were decreased; in 865 of 1000 cases, mortality was reduced by the new therapy, while in 135 it was increased).

Stinnett notes that the bootstrap method can also be used to correct for bias in the cost-effectiveness ratio that may be present when the ratio is based on the mean cost and effect estimates derived from relatively small samples *(86)*.

Sample-Size Analysis

Since the outcome of interest for economic analysis is the absolute magnitude of cost difference, it is appropriate to conduct sample size analysis from the perspective of gauging the likely precision of a CI *(87,88)*. The results of economic studies from clinical trials may also be used to compute cost-effectiveness or cost-utility ratios. It is important, then, that clinical trials with important economic components be designed so that the clinical benefit, cost or resource utilization, and cost-effectiveness ratios can be estimated reliably. This can be accomplished by determining the sample size that will yield acceptably small confidence limits for the three quantities of interest *(89)*. The limiting factor in most cases will be the precision of the cost-effectiveness ratio, as this ratio inherits noise from both numerator and denominator, and imprecision in the denominator will cause great fluctuations in the ratio *(81)*. Other factors to consider in sample size analysis are the magnitude of the cost difference that is considered economically meaningful, the distributional characteristics of cost data, and the efficiency of estimation methods used to analyze cost data. A naïve evaluation of cost data like those in Tables 5 and 6 might indicate that a few hundred subjects per arm are adequate to establish cost differences of a few thousand dollars with 95% confidence. But lower precision would be expected for a cost-effectiveness ratio, and precision would be expected to deteriorate because of nonnormality of the data and missing and censored data. Application of Fieller's Theorem may be useful in estimating sample sizes based on confidence intervals for cost-effectiveness ratios *(90)*.

Because trials tend to be underpowered for some or all economic outcomes, an alternative approach to demanding strict adherence to conventional hypothesis testing is to establish a sample size that is sufficient for establishing that the specified economic outcome is true with some probability. For example, one could establish the probability that the cost-effectiveness ratio of a treatment is less than $20 000 per life-year saved. Definitive probabilities of effects such as this can only be achieved when the Bayesian posterior distribution is sufficiently sharp or traditional confidence limits are sufficiently narrow *(91)*.

Example. Distributions of charges and reimbursements for cancer treatment can exhibit strong positive skew and, thus, cannot be assumed to be normally distributed. A similarly shaped distribution with positive skew is the log-normal. While the log-normal may only provide a first-order approximation to the true underlying distribution of charges, reimbursements, or costs, it can in some limited examples provide a useful method to estimate sample sizes needed to test for differences in cancer treatment costs between two samples.

To calculate sample size for tests of differences in costs, it is possible to convert everything so that calculations can be performed with normal distributions.[4] One usually generates a log-normal distribution by first generating the Zs (for some μ and σ) and then exponentiating to get $X = exp(Z)$. This is the reverse problem, in that the moments of the log-normally distributed Ys are given and it is desired to return to the Zs to conduct the power calculations.

The following simplifying assumptions are made: the costs have a two-parameter log-normal distribution, and $CV_X = 1$ (where CV_X is the coefficient of variation or the standard deviation of X divided by the mean of X).

We then solve following equality:

$$sqrt\{exp(2\sigma^2) - exp(\sigma^2)\} = exp(\sigma^2/2),$$

so that β can then be calculated as $\beta = E(X)/exp(\sigma^2/2)$, with the result that $\mu = \ln \beta$.

In Fig. 2, we fix Type I error level, α, as 0.05 (one-sided) and power, or $1-\beta$ as, 0.80. The expected difference between groups, Δ, is specified as a percentage of the mean. Note that these calculations do not depend on a particular value for the mean but rather depend on the ratio of the standard deviation relative to the mean and Δ as a function of the mean.

Price Adjustments

Unit prices change over time and vary by geographic location. When trial costs occur over a significant period of time and/or multiple locations, price adjusters should be used to allow costs to be expressed in a common "base." Even when trial costs are incurred over a short period of time and/or at a single location, adjustments should be made if costs are to be related to a different time or place (e.g., update of 1990 trial costs to 1997 costs or extrapolation of trial costs conducted in Olmstead County, Minnesota, to the United States overall).

Methods of adjusting prices are readily available *(92)* and include the medical component of the Consumer Price Index (CPI), which tends to overestimate overall inflation of health care costs, and the Fixed Weight Index (FWI) from HCFA, which results in lower estimates of inflation than the CPI. The FWI relies more on changes in the input prices of goods and services that go into the production of health care services. Two components of the FWI that may be of particular interest are the Prospective Payment System (PPS) input price index for hospital costs and the Medical Economic Index (MEI), which mea-

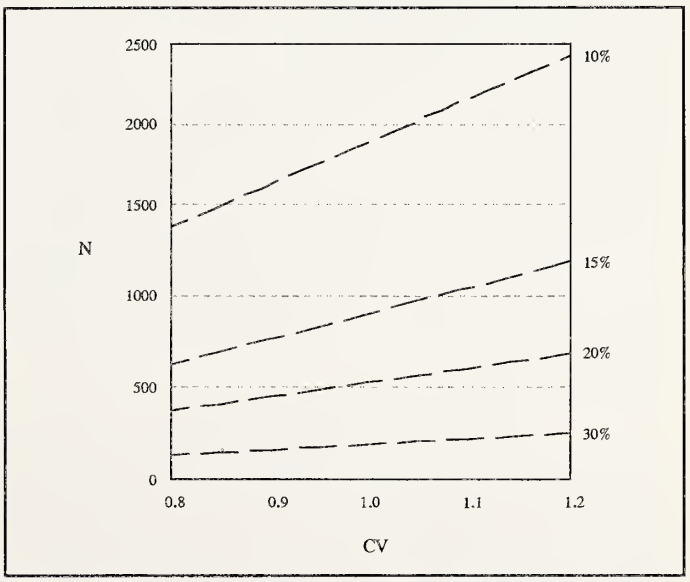


Fig. 2. A case study of sample size requirements. Sample size required for $\alpha = 0.05$ (one-sided), $1 - \beta = 0$ for comparing costs between two samples. N refers to the combined sample sizes of the two equally sized groups. CV refers to the coefficient of variation. Lines indicate sample size need for Δ% differences over ranges of CV. (Unpublished figure and data provided by Kimberly McGuigan and R. D. Fricker at the Rand Corporation.)

sures changes in the cost of physicians' time and operating expenses. Price adjustment has also been accomplished using the Geographic Adjustments Factors (GAF) for hospital wage costs and physician practice costs, including costs at the county level.[5]

Discounting the Costs and Benefits of Cancer Care

Economic theory indicates that assessments that look at costs and benefits of medical therapies over periods of time longer than a year should incorporate discounting. This concept embodies the fact that the costs and outcomes may occur at different points in time and that their value changes with time.

Costs incurred and effects realized in the present are generally greater than costs and effects of similar nominal value incurred later—if not used in the present, the resources could otherwise be invested in the interim. Hence, it is necessary to express all costs and benefits at one point in time, usually the present. This is done by "bringing back" all future costs and effects of a therapy and expressing them in terms of the present value.

Conclusion

Resource utilization and cost data collected within cancer clinical trials pose a series of important analytic challenges for economic investigators. These challenges stem from the need to analyze the data collected on a longitudinal basis within clinical trials as compared with dichotomous outcomes such as survival. Given the nature of data collected within economic studies, the potential for missing data is greatly increased over that usually seen in clinical trials without economic components. This section presents a state-of-the-art review of methodologic issues in economic evaluation of cancer clinical trials and highlights the analytic issues for investigators pursuing analysis of economic data in clinical trials. The section also services as an important reference to help all members of the economic evaluation team better understand the complexity of economic study design.

Part VI: Economic Evaluation in the Cooperative Group Setting

This section addresses practical issues that affect the ability of cooperative groups to collect data in multisite clinical trials. Because data will be gathered at different sites, clear and specific arrangements must be made from the beginning of the trial regarding data collection processes. It is recommended that discussion about an economic component—including how appropriate, complicated, and extensive it is—should be part of the concept review of a trial. The following issues are relevant to the design of an economic study in a clinical trial.

Are the Data Required?

If resource utilization or cost data are included in a trial, they should have the status of required clinical data and be part of any monitoring and reporting system. But should the commitment to collecting resource utilization or cost data be required for patient registrations? The best way to address this problem is to link therapeutic trial registration to completion of ancillary outcomes such as the economic study (i.e., submission of resource data is designated as a formal eligibility requirement in the protocol). Most resource-utilization data required for economic analysis are routinely collected for clinical outcomes. A resource-utilization strategy should not, therefore, be burdensome to clinical research associates (CRAs) or investigators and should not result in substantial biases in recruitment to the economic study. A separate consent for release of billing data may be required. If resource-utilization data are being collection, the patient can be informed in the main consent form that the medical charts also will be reviewed for cost outcomes.

However, even in a single cooperative group trial, some centers do not enter patients into every part of a trial. Some groups, for example, do not require that patients participate in ancillary studies (e.g., provide quality-of-life outcome data); a patient can still be eligible for the treatment component of the trial as long as the ancillary study is offered to the patient. This approach has the potential to bias the ancillary study outcomes, as physicians might only offer the quality-of-life study to a less ill subset of patients. Economic investigators are also concerned about this potential for bias. If it is not possible to include all patients, one compromise is to select *a priori* subsamples of patients; for example, include the first 500 patients randomized to the trial in the cost study (given that they meet the eligibility requirements). Data obtained with this approach can also be biased depending on historical cohort factors, such as the introduction of a new supportive care treatment midway through the trial. In this case, the cost-saving or cost-inducing effects will be missed.

Who Will Collect the Data at Cooperative Group Institutions?

CRAs collect and submit clinical data in most cooperative group settings. Will CRAs also collect the cost data, or should these data come directly from patient accounts offices at the institutions? It is also possible to collect cost data directly from patients. If collection of data from patients is necessary, this should be done by the local site and not the central data center (unless outside support is obtained for this effort, and then only with a communication linkage with the clinical investigator). Collection of data directly from patients by central office staff may also involve additional informed consent issues requiring local IRB approval. A combination of these data sources can be considered for cooperative group research.

How Will Data be Collected, Monitored, and Submitted?

The mechanism for data collection should be specified in the protocol. Part IV of this workbook presents a number of data collection sources for this effort. Forms represent the most common method used in cooperative group trials, and they can be mailed, faxed, or electronically submitted to the coordinating center. In addition, data can be submitted to the central office on diskette or tape. A number of procedural issues must be specified in the design of the economic study:

Highlight up front the importance of monitoring and data quality.

Set standards for data collection.

Include all resource-utilization and cost data in the group's system for monitoring data submission (e.g., "expectation systems"). This may require early input from programming staff.

To ensure that the necessary data are collected, be certain that protocols are written clearly and that terms and procedures are defined precisely.

Develop an analysis plan for the economic study.

Example. Detailed data collection procedures should be developed for each component of the protocol. An example of the issues to be addressed in developing these procedures follows for the collection of economic information from patients by telephone.

Who will be the contact person?

Is the contact person from the local hospital or a centralized location?

Are sufficient staff available?

Is there an informed consent issue? Does the clinical trials consent form require special permission for this data collection approach? Does it require the consent of the physician contact?

How will confidentiality be monitored? Are there additional IRB issues to consider in economic studies? Does clinical trial medical chart access obtained in consent forms for clinical data needs cover the level of information required for economic studies? At the very least, amend consent forms to include the cost objectives of the trial. If patient bills are evaluated, additional IRB approval may be required, depending on the local IRB. Speaking directly to IRBs concerning the financial use of the data especially in terms of confidentiality may assist in disclosure of needed information.

Do existing quality control procedures implemented in cooperative group trials address data quality issues for economic studies, particularly for billing or resource-utilization data collection from the patient? Economics investigators, coordinating center programmers, and statisticians should communicate early on about any particular features of economic data that may need to be incorporated into data entry verification systems and tracking systems. Procedures for standardizing the collection of all economic data are necessary but this is particularly important for data obtained for care delivered "off-site" (e.g., at other hospitals, outpatient clinics, nursing homes, in-home). Data that require additional documentation at the institution level (e.g., data obtained through a telephone interview) and allowable source documentation must be clearly specified.

Given these arrangements and approvals, the monitoring plan can incorporate early and ongoing review of the correctness of economic data as it is received at the coordinating center. This would allow changes to be made in the data collection process if appropriate data are not being collected. It is important to specify how long economic data will be collected and monitored.

Training: An Important Quality-Control Procedure

Training at cooperative group meetings is one way to achieve standardization in procedures for economic studies. Explanation of economic study objectives and definition of terms are important when new research outcomes are introduced in cooperative group trials. Roundtable sessions on various research topics are held for CRAs at cooperative group meetings, and these meetings could easily incorporate sessions on economic evaluation. These training sessions can also provide a vehicle for obtaining feedback on data collection forms for economic studies. Investigator meetings, group meetings, or nurse/CRA training sessions can also provide an opportunity for reviewing data collection forms. All of these meetings provide the opportunity to ask volunteer institutions to conduct pilot tests for the economic study forms and procedures. If possible, it is ideal to have separate training sessions for economic studies of cancer treatment protocols. However, time is very limited at these meetings, and sessions separate from the main training sessions may not be well attended. It is advisable to include at least limited training on economic studies during the regular training sessions at cooperative group meetings.

Consider terminology in training efforts. Confusion can arise around semantics in collecting quantities of resource use. It is important for the staff to recognize that they will be collecting familiar information; for example, numbers of procedures and surgeries. Spreadsheets, made up before the economic study is implemented, should list all major categories of data collection for economic studies. These can be used to explain the economic study to the CRAs. Probably 90% of the data for the economic study are already being collected for the clinical study—but in a different format. Few additional data items really need to be checked. Unit price information, beyond hospital bills, usually is not obtained at the institutional level. If patient bills are required for the study, they could be sent from the hospital finance center directly to the data center, given that appropriate arrangements have been agreed upon (e.g., confidentiality safeguards).

If resources allow, in-depth, on-site training can occur for CRAs at selected institutions (e.g., institutions where more detailed cost data might be collected). Cooperative group newsletters can provide another vehicle for training. It may be useful to add economic outcomes to the training provided for new study coordinators for cooperative group trials. A video tape is a useful approach for communicating economic study procedures and, if distributed to all participating institutions, can address staff turnover occurring between group meetings.

Small-Scale Pilot Studies

Small-scale pilot studies for data collection forms can be conducted to develop standard data collection forms for economic evaluations and to identify data collection problems. These small-scale pilot studies would address feasibility issues at the institutional level, as well as statistical center issues (e.g., integration into the current database system, extra programming requirements). Logs of time and effort required to use the forms and collect the data should be completed at institutions and statistical centers. Quantification of the extra effort involved, in terms of time and money (e.g., amount of time required to review charts, the source of the data required, and the cost in terms of staff time) can identify ways to reduce the data collection burden and to prepare budgets for future funding requests.

Several questions should be addressed in a pilot study. An assessment of the acceptability of the data collection forms can be done to ensure that the forms will be completed. The wording and structure of the questions in the forms must be assessed to ensure that the desired information is obtained.

Who Will Analyze and Share the Data?

Sharing of Data

For most cooperative groups, all data flow through a central statistical office. There are many safeguards to prevent the establishment of "shadow" databases. The clinical trials group may insist that all data remain at the statistical center, with output required by the economics group generated by the central office. However, this may cause economic investigators some

concern. With the economics expertise in a separate environment, biostatisticians at the statistical center probably do not have an analysis plan appropriate for economic outcomes or have access to the full extent of available related data. Even in this context, however, most statistical centers will prefer maintaining the database for the trial. One solution is to allow for substantial input from the economists regarding data analysis. In these collaborative relationships, therefore, communication between economic investigators and group biostatisticians should occur early in the design of a trial to address appropriate statistical methods for the cost data and the use of clinical data for the cost analyses.

Statistical center personnel often collaborate with outside economists, and some arrangements for data sharing may already have been discussed. Also, there may be a person who has the appropriate expertise for these analyses at the coordinating center, and there will be no need to split up the information. It may well be the case that the economic investigators will need all of the data to integrate resource consumption for patients with specific clinical practice outcomes. If these arrangements can be agreed upon, it is important to establish some mechanism to make sure that both groups are given continuous and consistent data. Before the clinical study begins, a decision should be made concerning the amount and type of data that will be given to the economic investigators and the publications rules for the economic study. Decisions need to be specified clearly and within the constraints established for that cooperative group.

If a decision is made to have economic data analyzed elsewhere, most cooperative groups will prefer to have all data sent to the statistical or coordinating center for forwarding to the economic investigators. If data are to be submitted directly to an economics center, the procedures for this separate submission should be clearly stated in the protocol.

Data Analysis Issues

Economic analysts should remain blinded until the last step in the analysis. This will avoid implicit conflict-of-interest issues and maintain the integrity of the overall trial. If data analysis involves the clinical data, the analysis ideally should not be done until the data are mature. Any use of major clinical endpoints will require waiting until data are mature. These decisions will be made by the data monitoring committee established for the particular trial or by arrangements established during the design phase of the trial.

Intergroup Collaboration Issues Specific to Cost Studies

It is generally unworkable to have one cooperative group execute and coordinate an ancillary study that is being coordinated by a different cooperative group because of the complications associated with transfer of large amounts of data. Therefore, it is reasonable to expect that cost studies will usually need to be executed and coordinated by the group that is coordinating the therapeutic trial. There are, of course, potential difficulties associated with this policy:

- negotiations of "ownership of economic data";
- no salary support for economists;
- noncoordinating group often does not receive appropriate credit (i.e., publications as primary author for economic study results); and
- awkward when the participating group may have more expertise in the area (quality of life, economics) than the coordinating group.

There are ways to encourage collaboration:

- the use of outside or separate funding (grants) to help support various areas of the trial;
- subcontracting of specific, external (to the group) staff to obtain required expertise;
- development of a liaison committee of economic experts among cooperative groups;
- attendance at each other's group meetings;
- individuals that focus on specific sites (i.e., breast or lung) can attend those committees at other cooperative group meetings to monitor upcoming trials that might be candidates for an economic study;
- the standardization of data collection instruments for economic protocols across groups, whenever possible; and
- it is recommended that discussion about an economic component including how appropriate, complicated, and extensive it is should be part of the concept review of an intergroup trial.

Quality Control

Quality control issues become more challenging in the intergroup trial context because of the variation in procedures employed by the different groups. Quality-control procedures suggested for single group trials are equally relevant to the intergroup context. In particular, training at the meetings held by each group would be very important to standardize data collection procedures. Regular communication between intergroup staff associated with the cost study would help identify problems with forms and procedures and allow for improvements in procedures.

Intergroup Pilot Test

Clinical and economic investigators across cooperative groups interested in a specific protocol could identify clinical research areas with important economic issues and identify trials planned in those areas; the cooperation of the Cooperative Group Chairs Committee would assist this process. Pilot work is particularly important for intergroup trials, given existing variations in procedures. Each group could evaluate the acceptability of the data collection forms, particularly with respect to access to required information and the time required to collect the information. Quantification of the extra effort involved, in terms of time and money (e.g., amount of time required to review charts, the source of the data required, and the cost in terms of staff time) could document whether or not intergroup economic studies were feasible. The intergroup pilot study could identify ways to reduce the data collection burden and to prepare budgets for future funding requests.

Data Issues

Information-Sharing

The same questions addressed for single-group economic studies involving outside economic expertise apply to intergroup studies. Who will have access to cost (and clinical) data? What

is the process for sharing data, and how will the information be copied and distributed to others? Which data are relevant to the analysis of economic and clinical parameters?

How Will Economic Studies Be Funded in Cooperative Group Trials?

The conduct of economic studies requires expenditure of considerable staff resources, both at the coordinating center and at individual institutions. Because these studies are new, case report forms, data collection procedures, data base preparation, and data analysis all require considerable start-up effort. Economic outcome data must be accorded the same importance as clinical data, meaning that, if collected, they are subject to the same strict quality control procedures that drive collection of clinical endpoint data. Funding is required to support this effort. Currently, we recognize the importance of including economic outcomes in selected clinical trials. Introduction of quality-of-life outcomes to cooperative groups is a case in point, where substantial central staff time for quality control and monitoring has been necessary *(93)*. However, identifying funding sources

Appendix I. Excerpt from Clinical Trial Case Report Form

SOUTHWEST ONCOLOGY GROUP PROTOCOL S9509
A Randomized Phase III Trial of Paclitaxel Plus Carboplatin Versus Vinorelbine and Cisplatin in Untreated Non–Small-Cell Lung Cancer

SPECIAL INSTRUCTIONS: RESOURCE UTILIZATION SUMMARY FORM
The purpose of tracking resource utilization is to see if patients on one treatment arm require more medical resources than those on the other arm (i.e., to see if the treatment arms differ with respect to cost of care). We will be tracking medical resource use more intensively during the first 25 weeks postrandomization. We will monitor medical care between months 6 and 24, but less intensively. We have targeted medical resources that are particularly expensive, and for that reason, many commonly used tests or treatments may not be listed on the forms. Completion of the summary forms requires chart review. It may be necessary to consult physician progress notes, physician orders, and nursing notes. Since chart review is normally done at the end of each treatment cycle, we suggest that you collect information for the Resource Utilization Summary Forms each cycle. Funding has been provided for the extra chart review time required for the medical resource use study. If you are in a managed care setting, you may find it easier to obtain this information from the billing office.

II. SUPPORTIVE CARE MEDICATIONS/TREATMENT

Were supportive care medications/treatment utilized during this period? ☐ No ☐ Yes

	Number of Days Medications Ordered for:					
	Inpatient		Outpatient		Home Care	
	Infusion	Noninfusion	Infusion	Noninfusion	Infusion	Noninfusion
Antiemetics: 5HT3 Antagonists: (ondansetron, granisetron)						
Parenteral Non-5HT3 Agents: (metoclopramide, prochlorperazine, haloperidol, diphen-hydramide hydrochloride, dexmethasone)						
Propulsid (cisapride)						
Benzodizepines (lorazepam)						
Bone Pain Therapy: Bisphosphonates: (etidronate, pamidronate)						
Strontium						
Narcotics (epidurals, pumps)						
Narcotic skin patches						
Antibiotics (infusion only)						
Growth Factors: (G-CSF/filgrastim, GM-CSF/sargramostim)						
(erythropoetin)						

that fit the specific timing and procedural constraints of cooperative group research is a challenge.

To date, several approaches to funding economic studies have been tried in cooperative group trials. Pharmaceutical industry funding has supported several economic studies in the cooperative groups: CALGB 9411 (an acute lymphocyte leukemia trial) and SWOG 9509 (an advanced-stage non-small-cell lung cancer trial) are examples.

Third-party payer examples are less common, but one currently in place supports an intergroup economics investigation in a metastatic breast cancer trial by U.S. Healthcare. The extent to which other large managed care organizations can be convinced to support these studies is yet to be determined.

The NCI has two potential mechanisms for supporting economic studies: direct grants, and credits based on accrual to either treatment or cancer control trials. Through a request for applications, NCI is supporting trials of minimally invasive surgery that incorporate an economic component. An example is an intergroup trial (INT 0146) of open colectomy versus laparoscopic-assisted colectomy for colon cancer. Pursuing the RO1 mechanism involves delays between the time applications are submitted and funds are awarded. The cooperative group will not be able to delay activation of a trial while funding for the economic component is being sought.

Cancer control and treatment credits represent another potential avenue for funding. Treatment credits translate into data management support based on accrual effort during the previous grant cycle. Cancer control credits are assigned to cancer prevention and control clinical trials following review and approval by the Division of Cancer Prevention (DCP) Cancer Control Protocol Review Committee (CCPRC). The process is analogous to the review of treatment protocols in the Cancer Treatment Evaluation Program.

Credit is a measure of the data management burden needed to implement an intervention in a clinical trial. Most treatment clinical trials receive one credit/accrual. Credits for cancer prevention and control clinical trials vary depending on the nature of the intervention, the amount of data collection, and the length of follow-up. Community Clinical Oncology Programs (CCOPs) receive budgets based in part on their projected accrual as measured in treatment and cancer control credits. Cooperative groups serving as CCOP research bases receive "pass-through" finding for members and affiliates that accrue to cancer control clinical trials.

To date, economic evaluation in cooperative group treatment studies have not received cancer control credits. Quality-of-life studies have received cancer control credit when they were stand-alone protocols or when the outcomes have been incorporated into treatment trials. Economic outcomes could follow the same model, but this will require a reassessment of DCPC policy. Another possibility is award of additional treatment credits for registrations to companion economic trials or treatment trials with an economic outcome.

As can be seen by this brief summary, routine funding mechanisms are not available to support economic studies in cooperative group trials. Access to stable funding for this research is a first step in our efforts to encourage cooperative groups to incorporate economic outcomes data in cancer elinical trials.

This section addresses specific issues related to the implementation of economic evaluations within the NCI cooperative group setting. The NCI cooperative groups are unique and specialized research institutions with well-established protocols for collection of clinical data. However, economic evaluation is a new endpoint within cancer clinical trials for which investigators may have to modify their study procedures. Given the need for change, training and communication become key factors for success in the implementation of economic evaluation in the cooperative group setting. Finally, given the current structure of funding for NCI clinical trials, economic evaluation often will require external funding arrangements, including peer-reviewed grants or industry support.

Conclusion

Economic evaluation of cancer clinical trials is a new and evolving discipline. As outlined in this workbook, implementation of economic evaluation requires a multidisclipinary research team and extensive planning and forethought analogous to the process of designing a study to assess the clinical endpoints of a clinical trial. Economic evaluation requires training of cooperative group personnel as well as training of site personnel including physicians and the clinical research study team. Through this collaboration, economic data can be successfully collected within cancer clinical trials.

References

(1) Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.

(2) Drummond MF, O'Brien B, Stoddart GL, Torrance DW. Methods for the economic evaluation of health care programmes: second edition. New York: Oxford University Press.

(3) Parker SL, Tong T, Bolden S, Wingo PA. Cancer statistics, 1997. Cancer 1997;47:5–27.

(4) Brown ML. The national economic burden of cancer. J Natl Cancer Inst 1990;82:1811–4.

(5) Rundle RL. Salick pioneers selling cancer care to HMOs. The Wall Street Journal 1996;August 12, 1996, B1–B2.

(6) Brown ML, Hodgson TA, Rice DP. Economic impact of cancer in the U.S. In: Schottenfeld D, Fraumeni J, editors. Cancer, Epidemiology, and Prevention. New York: Oxford University Press, 1996.

(7) Callahan D. Controlling the costs of health care for the elderly: fair means and foul. N Engl J Med 1996;335:744–6.

(8) Levinsky NG. The purpose of advance medical planning: autonomy for patients or limitation of care? N Engl J Med 1996;335:741–3.

(9) Hillner BE. Potential evaluation of the incremental cost-effectiveness of paclitaxel in advanced non-small-cell lung cancer (Eastern Cooperative Oncology Group 5592). Monogr Natl Cancer Inst 1995;19:65–7.

(10) American Society of Clinical Oncology. Recommended breast cancer surveillance guidelines. ASCO Breast Cancer Surveillance Expert Panel. J Clin Oncol 1997;15:2149–56.

(11) American Society of Clinical Oncology. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. ASCO Tumor Marker Expert Panel. J Clin Oncol 1996;14:2843–77.

(12) American Society of Clinical Oncology. Recommendations for the use of hematopoietic CSFs: evidence-based clinical practice guidelines. ASCO Ad Hoc Colony-Stimulating Factor Guidelines Expert Panel. J Clin Oncol 1994;12:2471–508.

(13) Peters WP, Ross M, Vredenburgh JJ, Hussein A, Rubin P, Dukelow K, et al. The use of intensive clinic support to permit outpatient autologous bone marrow transplantation for breast cancer. Semin Oncol 1994;21(4 suppl 7):25–31.

(14) Bennett CL, Westerman IL. Economic analysis during phase III clinical trials: who, what, when, where, and why? Oncology 1995;9(suppl):1–7.

(15) Hillner BE, Kirkwood JM, Atkins MB, Atkins ER, Johnson ER, Smith TJ. Economic analysis of adjuvant interferon alfa-2b in high-risk melanoma based on projections from Eastern Cooperative Oncology Group 1684. J Clin Oncol 1995;15:2351–8.

(16) Bochner F, Martin ED, Burgess ND, Somogyi AA, Misan GM. How can

hospitals ration drugs? Drug rationing in a teaching hospital: a method to assign priorities. Br Med J 1994;308:901–5.
(17) Smith TJ, Bodurtha JN. Ethics in oncology: balancing interests of the patient, oncologist, and society. J Clin Oncol 1995;13:2464–70.
(18) Eisenberg JM. Clinical economics: a guide to the economic analysis of clinical practices. JAMA 1989;262:2879–86.
(19) Bennett CL, Matchar D, McCrory D, McLeod DG, Crawford ED, Hillner BE. Cost-effective models for flutamide for prostate carcinoma patients: are they helpful to policy makers? Cancer 1996;77:1854–61.
(20) Hillner BE, Smith TJ, Desch CE. Efficacy and cost-effectiveness of autologous bone marrow transplantation in metastatic breast cancer: estimates using decision analysis while awaiting clinical trial results. JAMA 1992; 267:2055–62.
(21) Davies E, Clarke C, Hopkins A. Malignant cerebral glioma—I. Survival, disability, and morbidity after radiotherapy. Br Med J 1996;313:1507–12.
(22) Davies E, Clarke C, Hopkins A. Malignant cerebral glioma—II. Perspectives of patients and relatives on the value of radiotherapy. Br Med J 1996;313:1512–6.
(23) Drummond MF. Survey of cost-effectiveness and cost-benefit analyses in industrialized countries. World Health Statistics Quarterly 1985;38: 383–401.
(24) Drummond MF, O'Brien B. Economic analysis alongside clinical trials: practical considerations. The Economics Workgroup. J Rheumatol 1995; 22:1418–9.
(25) Schulman KA, Dorsainvil D, Yabroff KR, DiCesare J, Feltzer J, Glick HA, et al. Prospective economic evaluation of rhGM-CSF/rhIL-3 bone marrow transplant for leukemia and lymphoma (manuscript in preparation).
(26) Lerman C, Narod S, Schulman KA, Hughes C, Gomez-Caminero A, Bonney G, et al. BRCA1 testing in families with hereditary breast-ovarian cancer: a prospective study of patient decision-making and outcomes. JAMA 1996;275:1885–92.
(27) Haioun C, Lepage E, Gisselbrecht C, Coiffier B, Bosly A, Tilly H, et al. Comparison of autologous bone marrow transplantation with sequential chemotherapy for intermediate-grade and high-grade non-Hodgkin's lymphoma in first complete remission: a study of 464 patients. J Clin Oncol 1994;12:2543–51.
(28) Peters WP, Ross M, Vredenburgh JJ, Meisenberg B, Marks LB, Winer E, et al. High dose chemotherapy and autologous bone marrow support as consolidation after standard-dose adjuvant therapy for high-risk primary breast cancer. J Clin Oncol 1993;11:1132–43.
(29) Feuer EJ, Frey CM, Brawley OW, Nayfield SG, Cunningham JB, Geller NL, et al. After a treatment breakthrough: a comparison of trial and population-based data for advanced testicular cancer. J Clin Oncol 1994;12: 368–77.
(30) Goldhirsch A, Wood WC, Senn HJ, Glick JH, Gelber RD. Meeting highlights: international consensus panel on the treatment of primary breast cancer. J Natl Cancer Inst 1995;87:1441–5.
(31) Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. Br Med J 1996;313:275–83.
(32) Luce BR, Manning WG, Siegel JE, Lipscomb J. Estimating costs in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
(33) Schulman KA, Glick HA. Economic evaluation of high dose chemotherapy and BMT for patients with stage 2 and stage 3 breast cancer with greater than ten positive nodes at the time of surgical excision. J Natl Cancer Inst 1995;19:45–50.
(34) Olivotto IA, Bajdik CD, Plenderleith IH, Coppin CM, Gelmon KA, Jackson SM, et al. Adjuvant systemic therapy and survival after breast cancer. N Engl J Med 1994;330:805–10
(35) Mauskopf JK, Schulman K, Bell L, Glick H. A strategy for collecting pharmacoeconomic data during phase II/III clinical trials. PharmacoEconomics 1996;9:264–77.
(36) Kukull WA, Koepsell TB, Conrad DA, Immanuel V, Prodzinski J, Franz C. Rapid estimation of hospitalization charges from a brief medical record review: evaluation of a multivariate prediction model. Med Care 1986;24: 961–6.
(37) Nefcy P. Challenge for health care: information for managing costs, utilization and quality. In: QQPHS 1992 Conference Proceedings. Atlanta: Institute of Industrial Engineers, 1992.
(38) Hsiao WC, Braun P, Dunn D, Becker ER. Resource-based relative values: an overview. JAMA 1988;260:2347–53
(39) Federal Register. FR 58. Washington (DC): National Archives and Records Administration, 1993.
(40) Bennett CL, Cvitanic M, Pascal A. The costs of AIDS in Los Angeles. J Acq Immune Defic Syndr 1991;4:197–203.
(41) Edwards WS, Berlin M. National Medical Expenditure Survey: Questionnaires and Data Collection Methods for the Household Survey and the Survey of American Indians and Alaska Natives, Methods 2. DHHS Publ No. (PHS) 89–3450. Rockville (MD): DHHS, 1989.
(42) Finkler S. The distinction between cost and charges. Ann Intern Med 1982;96:102–9.
(43) Finkler SA. Cost Accounting for Health Care Organizations: Concepts and Applications. Gaithersburg (MD): Aspen Publishers, 1994.
(44) UB-92 Editor: A Guide to Medicare Billing. Falconer C, editor. Reston (VA): St. Anthony, 1994.
(45) Red Book 1995. Drug topics. Montvale (NJ): Medical Economics Data, 1995.
(46) Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG. Potential for cancer related health services research using a linked Medicare-tumor registry database. Med Care 1993;31:732–48.
(47) Coyle D. Statistical analysis in pharmacoeconomic studies: a review of current issues and standards. PharmacoEconomics 1996;9:506–16.
(48) Dudley RA, Harrell FE Jr, Smith LR, Mark DB, Califf RM, Pryor DB, et al. Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. J Clin Epidemiol 1993;46:261–71.
(49) Zhou XH, Gao S. Confidence intervals for log-normal mean. Stat Med 1997;16:783–90.
(50) Zhou XH, Gao S, Hui SL. Methods for comparing the means of two independent log-normal samples. Biometrics. In press.
(51) Edgington ES. Randomization Tests. New York: Marcel Dekker, 1986.
(52) Efron B, Tibshirani RJ. An Introduction to Bootstrap. New York: Chapman and Hall, 1993.
(53) Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. Inter Stat Rev 1993;59:227–40.
(54) Schoenfeld DA. Sample size formulae for the proportional hazards regression model. Biometrics 1983;39:499–503.
(55) Anderson GL, Fleming TR. Model misspecification in proportional hazards regression. Biometrika 1995;82:527–41.
(56) Ford IJ, Norrie J, Ahmadi S. Model inconsistency, illustrated by the Cox proportional hazards model. Stat Med 1995;14:735–46.
(57) Akazawa KT, Nakamura T, Palesch Y. Power of log-rank test and Cox regression model in clinical trials with heterogeneous samples. Stat Med 1997;16:583–97.
(58) Rutten-van Moken MP, van Doorslaer EK, van Vliet RC. Statistical analysis of cost outcomes in a randomized controlled clinical trial. Health Economics 1994;3:333–45.
(59) Duan N. Smearing estimate: a nonparametric retransformation method. J Am Stat Assoc 1983;78:605–10.
(60) Cox DR, Oakes D. Analysis of survival data. London: Chapman and Hall, 1984.
(61) Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. J Am Stat Assoc 1989;84:1074–8.
(62) Fenn P, McGuire A, Backhouse M, Jones D. Modeling programme costs in economic evaluation. J Health Economics 1996;15:115–25.
(63) Smith TJ, Desch CE, Hillner BE. Ways to reduce the cost of oncology care without compromising the quality. Cancer Invest 1994;12:257–64.
(64) Harrell FE Jr, Lee KL, DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87.
(65) Etzioni RD, Urban N, Baker M. Estimating the costs attributable to a disease with application to ovarian cancer. J Clin Epidemiol 1996;49: 95–103.
(66) Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. Biometrics. In press.
(67) Riley GF, Potosky AL, Lubitz JD, Kessler LG. Medicare payments from diagnosis to death for elderly cancer patients by stage at diagnosis. Med Care 1995;33:828–41.
(68) Etzioni RD, Feuer RJ, Sullivan SD, Lin D, Hu C, Ramsey S. On the use of survival analysis techniques to estimate medical care costs. Working Paper, Fred Hutchison Cancer Research Center, Seattle, 1997.
(69) Cole BF, Gelber RD, Anderson KM. Parametric approaches to quality-adjusted survival analysis. Biometrics 1994;50:621–31.
(70) Cole BF, Gelber RD, Kirkwood JM, Goldhirsch A, Barylak E, Borden E. Quality-of-life–adjusted survival analysis of interferon alfa-2b adjuvant treatment of high-risk resected cutaneous melanoma: an Eastern Cooperative Oncology Group study. J Clin Oncol 1996;14:2666–73.
(71) Draper D. Assessment and propagation of model uncertainty. J R Stat Soc 1995;57:45–97.
(72) Little T, Rubin D. Statistical analysis with missing data. New York: John Wiley & Sons, 1987.
(73) Rubin DB. Multiple imputations for nonresponse in surveys. New Work: John Wiley & Sons, 1987.
(74) Rubin D, Shenker N. Multiple imputation in health-care data bases: an overview and some applications. Stat Med 1991;10:585–98.

(75) Lavori P, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. Stat Med 1995;14:1913–25.
(76) Efron B. Missing data, imputation, and the bootstrap. J Am Stat Assoc 1994;89:463–79.
(77) McGuigan KA, Ellickson PE, Hays RD, Bell RM. Adjusting for attrition in school-based samples: bias, precision, and cost trade-offs of three methods. Eval Rev 1997;21:554–67.
(78) Karrison TG. Restricted mean life with adjustment for covariates. J Am Stat Assoc 1987;82:1169–76.
(79) Karrison TG. Use of Irwin's restricted mean as an index for comparing survival in different treatment groups—Interpretation and power considerations. Controlled Clinical Trials 1997;18:151–67.
(80) Mark DB, Hlatky MA, Califf RM, Naylor CD, Lee KL, Armstrong PW, et al. Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. N Engl J Med 1995;332:1418–24.
(81) Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In: Gold MR, Russell LB, Siegel JA, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.
(82) Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. Health Economics 1997;6:243–52.
(83) Willan AR, O'Brien BJ. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. Health Economics 1996;5:297–305.
(84) Hlatky MA, Boothroyd DB, Johnstone IM, Marcus FI, Hahn EH, Hartz V, et al. Long-term cost-effectiveness of alternative management strategies for patients with life-threatening ventricular arrhythmias. J Clin Epidemiol 1997;50:185–93.
(85) Chaudhary X, Mohammad A, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. Stat Med 1996;15:1447–58.
(86) Stinnett AA. Adjusting for bias in C/E ratio estimates. Health Economics 1996;5:470–2.
(87) Hornberger JC, Brown BW, Halpern J Jr. Designing a cost-effective clinical trial. Stat Med 1995;14:2249–59
(88) Borenstein, M. The case for confidence intervals in controlled trials. Controlled Clin Trials 1994;15:411–28.
(89) Borenstein M. Planning for precision in survival studies. J Clin Epidemiol 1994;47:1277–85.
(90) Willan AR, O'Brien BJ. Confidence intervals for cost-effectiveness ratios: an application of Fieller's Theorem. Health Economics 1996;5:297–305.
(91) Spiegelhalter DJ, Freedman LS, Parmar MK. Bayesian approaches to randomized trials. J R Stat Soc 1994;157:357–416.
(92) Sensenig AL, Heffler SK, Donham CD. Health care indicators: hospital, employment, and price indicators for the health care industry—third quarter 1995. Health Care Financing Review 1996;17:269–79.
(93) Moinpour CM. Costs of quality-of-life research in Southwest Oncology Group trials. J Natl Cancer Inst 1996;20:11–6.
(94) Smith TJ, Desch CE, Hillner BE. Efficacy and cost-effectiveness in oncology: rational allocation of cancer care resources. J Natl Cancer Inst 1996;85:1460–71.
(95) Smith TJ, Hillner BE, Schmitz N, Linch DC, Dreger P, Goldstone AH, et al. Economic analysis of a randomized clinical trial to compare filgrastim-mobilized peripheral blood progenitor cell transplantation and autologous bone marrow transplantation in patients with Hodgkin's and non-Hodgkin's lymphoma. J Clin Oncol 1997;15:5–10.
(96) Elit LM, Gafni A, Levine MN. Economic and policy implications of adopting paclitaxel as first-line therapy for advanced ovarian cancer: an Ontario perspective. J Clin Oncol 1997;15:632–9.

Notes

[1]*Editor's note:* SEER is a set of geographically defined, population-based central tumor registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Each registry annually submits its cases to the NCI on a computer tape. These computer tapes are then edited by the NCI and made available for analysis.

[2]These indices are available from Carolyn S. Donham, Office of the Actuary, Health Care Financing Administration, telephone 410–786–7947.

[3]For example, the absolute increase in 5-year survival probability, when a proportional hazards model holds, is $S^r - S$. Here, S is the 5-year survival for control subjects, and r is the experimental control hazard ratio. This absolute survival difference is strongly influenced by S. If the survival distribution for control patients is exponential with a constant hazard rate h and proportional hazards model holds, the difference in mean life length due to treatment is $\mu(1-r)/r$, where μ is the control group's mean life length $(1/h)$. When one wishes to estimate absolute treatment effects for patients different from those enrolled, these effects will depend on prognostic factors.

[4]The more general case is a three-parameter log-normal distribution, $Y \sim LN(\alpha, \beta, \sigma)$, with location parameter α, scale parameter β, and shape parameter σ. The expectation is $E(Y) = \alpha + \beta \exp(\sigma^2/2)$, and the variance, $var(Y) = \beta^2 exp(\sigma^2) [exp(\sigma^2) - 1]$. For the two-parameter log-normal distribution, the location parameter is assumed to be 0.

[5]In limited experience thus far, the proportional hazards assumption has been remarkably well satisfied. A case study of the Cox model in estimating mean and median costs along with methods for checking model assumptions may be obtained from the World Wide Web: http://www.med.virginia.edu/medicine/clinical/hes/heshp4.htm

We thank the many people who participated in this project, including the writing committee, the workshop participants, and outside reviewers at the NCI and other institutions across the eountry. We also thank Joy Lopez for her assistance in coordinating this project and the Amgen Corporation for its generous support of this activity.

If you're not already subscribing, here's what you're missing...

The *Journal of the National Cancer Institute* remains the Number 1 journal in the field (*ISI Journal Citation Reports*, October 1997), publishing the best original research papers in oncology from around the world. Internationally acclaimed as the source for the most up-to-date news and information from the rapidly changing fields of cancer research and treatment, the *Journal of the National Cancer Institute* is now available by subscription from Oxford University Press.

Each issue is packed with extensive news about important cancer-related science, policy, politics, and people. Plus, editorials and reviews by experts in the field, book reviews, and commentary on timely topics.

Become a subscriber, today.
www.oup.co.uk/jnci/

And there's more!

Immediately upon subscribing, you will become a member of the NCI Information Associates Program, which includes these outstanding benefits:

- One copy of each new *Journal of the National Cancer Institute Monographs**
- 24-hour access to PDQ®, the NCI's comprehensive cancer information database
- Electronic Bulletin Board System
- *ProtoCall*™ Newsletter
- Trained Member Service Representatives

NATIONAL CANCER INSTITUTE INFORMATION ASSOCIATES PROGRAM

For more information about the Information Associates Program, call 1-800-624-7890 (+301-496-7600 outside the USA).

*Free to members upon request prior to publication.

I wish to subscribe to the *Journal of the National Cancer Institute.*

Name (please print)_____

Address _____

_____

Zip/Post code_____ Country_____

❒ I enclosed the correct amount (check payable to Oxford University Press).

❒ Please bill me.

Please charge my

❒ MasterCard ❒ Visa ❒ American Express ❒ Diners ❒ JCB card.

Card number

Exp. date_____Signature _____

1998 Subscription Rates Volume 90 (24 issues)

*Institutional subscribers will now receive all monographs automatically as they are published with no need to order separately. If, however, you would prefer not to receive this automatic service, please contact your nearest *Journal* Subscriptions Department at Oxford University Press. Personal subscribers will have the opportunity to order monographs free of charge prior to publication.

Individuals	
US & Canada	❒ One year: US$132
UK & Europe	❒ One year: £127
Elsewhere	❒ One year: US$180
Institutions	
US & Canada	❒ One year: US$190*
UK & Europe	❒ One year: £170*
Elsewhere	❒ One year: US$240*

Single issues are available upon request. Air-expedited delivery fee is included for subscriptions outside North America. Individual subscriptions must be paid for by personal check or credit card and sent to private addresses only. Canadian residents add 7% GST.

To begin your subscription, call 1-800-852-7323 (USA & Canada). Or call 919-677-0977. Fax: 919-677-1714.
Outside the Americas, ring +44 (0)1865 267907 or fax +44 (0)1865 267485.

In the Americas, please return to:
Oxford University Press, Journals Marketing (f), 2001 Evans Road, Cary, NC 27513, USA.
Elsewhere, please return to:
Oxford University Press, Journals Marketing (f), Great Clarendon St., Oxford, OX2 6DP, UK.

OXFORD JOURNALS

NIH LIBRARY
3 1496 00759 9338

MONOGRAPHS
JOURNAL OF THE NATIONAL CANCER INSTITUTE

Cancer Risk Communication: What We Know and What We Need to Learn

1999
Number 25

Contents

Foreword 1
Richard D. Klausner

The Cancer Risk Communication Meeting in Perspective 3
Barbara K. Rimer

Workshop Keynote Address 5
The editors

Introduction of Section: Challenges Inherent in Communicating Cancer Risk Information 5
J. Paul Van Nevel

- **Why (Cancer) Risk Communication Can Be Hard** 7
  Baruch Fischhoff
- **Improving Cancer Risk Communication: a Discussion of Fischhoff** 14
  Edward Maibach
- **What Does It Mean to Understand a Risk? Evaluating Risk Comprehension** 15
  Neil D. Weinstein

Introduction of Section: Overarching Considerations in Risk Communications: Romancing the Message 21
Joann E. Rodgers

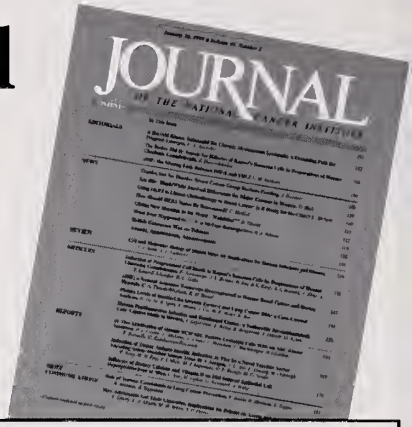
- **Communicating Health Risk to Ethnic Groups: Reaching Hispanics as a Case Study** 23
  Elmer E. Huerta, Everly Macario
- **Dealing With Competing and Conflicting Risks in Cancer Communication** 27
  Matthew W. Kreuter
- **New Directions for Risk Communication Research: a Discussion With Additional Suggestions** 35
  Alfred C. Marcus

Introduction of Section: Enabling Informed Decisions About Cancer Risk 43
Barbara Bowles Biesecker

- **Treating People With Information: an Analysis and Review of Approaches to Communicating Health Risk Information** 44
  Alexander J. Rothman, Marc T. Kiviniemi
- **Cancer Screening Decisions** 52
  Kevin D. McCaul, Heather E. Tulloch
- **Risk Communication in Genetic Testing for Cancer Susceptibility** 59
  Robert T. Croyle, Caryn Lerman
- **Decision Aids for Patients Considering Options Affecting Cancer Outcomes: Evidence of Efficacy and Policy Implications** 67
  Annette M. O'Connor, Valerie Fiset, Catherine DeGrasse, Ian D. Graham, William Evans, Dawn Stacey, Andreas Laupacis, Peter Tugwell
- **Population Risk, Actual Risk, Perceived Risk, and Cancer Control: a Discussion** 81
  Howard Leventhal, Kim Kelly, Elaine A. Leventhal

(Contents continued on back cover)

If you're not already subscribing to the #1 oncology journal in the field, here's what you're missing...



The *Journal of the National Cancer Institute* is now the Number 1 journal in its field (*ISI Journal Citation Reports*, October 1998), publishing the best papers in oncology from around the world. Internationally acclaimed as the source for the most up-to-date information from the rapidly changing fields of cancer research and treatment, the *Journal of the National Cancer Institute* is available by subscription from Oxford University Press.

Each issue is also packed with extensive news about important cancer-related science, policy, politics, and people. Plus, editorials and reviews by experts in the field, book reviews, and commentary on timely topics.

Become a subscriber, today.
www.oup.co.uk/jnci/

And there's more!

Immediately upon subscribing, you will become a member of the NCI Information Associates Program, which includes these outstanding benefits:

- One copy of each new *Journal of the National Cancer Institute Monographs**
- 24-hour access to PDQ®, the NCI's comprehensive cancer information database
- Listserv Subscription Service
- *ProtoCall*™ Newsletter
- Member Help Line Services

NATIONAL CANCER INSTITUTE
INFORMATION ASSOCIATES PROGRAM

For more information about the Information Associates Program, call 1-800-624-7890 (+301-496-7600 outside the USA).

*Free to members upon request prior to publication.

I wish to subscribe to the *Journal of the National Cancer Institute.*

Name (please print)_____

Address _____

_____

ZIP/Post code_____ Country_____

❒ I enclosed the correct amount (check payable to Oxford University Press).

❒ Please bill me.

Please charge my

❒ MasterCard ❒ Visa ❒ American Express ❒ Diners ❒ JCB card.

Card number

Exp. date_____Signature _____

1999 Subscription Rates Volume 91 (24 issues)

*Institutional subscribers will now receive all monographs automatically as they are published with no need to order separately. If, however, you would prefer not to receive this automatic service, please contact your nearest *Journal* Subscriptions Department at Oxford University Press. Personal subscribers will have the opportunity to order monographs free of charge prior to publication.

Individuals	
US & Canada	❒ One year: US$145
UK & Europe	❒ One year: £140
Elsewhere	❒ One year: US$198
Institutions	
US & Canada	❒ One year: US$210*
UK & Europe	❒ One year: £187*
Elsewhere	❒ One year: US$265*

Single issues are available upon request. Air-expedited delivery fee is included for subscriptions outside North America. Individual subscriptions must be paid for by personal check or credit card and sent to private addresses only. Canadian residents add 7% GST.

To begin your subscription, call 1-800-852-7323 (USA & Canada). Or call 919-677-0977. Fax: 919-677-1714.
Outside the Americas, ring +44 (0)1865 267907 or fax +44 (0)1865 267485.

In the Americas, please return to:
Oxford University Press, Journals Marketing (q), 2001 Evans Road, Cary, NC 27513, USA.
Elsewhere, please return to:
Oxford University Press, Journals Marketing (q), Great Clarendon St., Oxford, OX2 6DP, UK.

OXFORD UNIVERSITY PRESS

MONOGRAPHS

JOURNAL OF THE NATIONAL CANCER INSTITUTE

Number 25 *1999*
ISSN 0027-8874
ISBN 0-19-922476-5

EDITORIAL BOARD

Barnett S. Kramer
Editor-in-Chief

J. Gordon McVie
European Editor

Eric J. Seifter
Book Reviews Editor

J. Paul Van Nevel
News Editor

Frederic J. Kaye
Douglas L. Weed
Reviews Editors

Martin L. Brown
Economics Editor

ASSOCIATE EDITORS

Susan G. Arbuck
Frank M. Balis
William J. Blot
Peter M. Blumberg
John D. Boice, Jr.
Louise A. Brinton
Ross C. Donehower
Susan S. Ellenberg
Ellen G. Feigal
Suzanne W. Fletcher
Michael A. Friedman
Patricia A. Ganz
Edward L. Giovannucci
John K. Gohagan
Frank J. Gonzalez
Michael M. Gottesman
Peter Greenwald
Jean L. Grem
Donald E. Henson
Susan M. Hubbard
Colin R. Jefcoate
Arnold D. Kaluzny
Frederic J. Kaye
Hynda K. Kleinman
Theodore S. Lawrence
Bernard Levin
W. Marston Linehan
Marc E. Lippman
Scott M. Lippman
Darrell T. Liu
Dan L. Longo
Reuben Lotan
Douglas R. Lowy
Susan G. Nayfield
David L. Nelson
David G. Poplack
Ross L. Prentice
Alan S. Rabson
Edward A. Sausville
Robert H. Shoemaker
David Sidransky
Richard M. Simon
Michael B. Sporn
Maryalice Stetler-Stevenson
J. Paul Van Nevel
Douglas L. Weed
Noel S. Weiss

STATISTICAL EDITORS

Janet W. Andersen
Jacques Benichou
Donald A. Berry
Barry W. Brown
Bernard F. Cole
Susan S. Ellenberg
Scott S. Emerson
Eric Feuer
Mary Foulkes
Mitchell H. Gail
Edmund A. Gehan
Barry I. Graubard
Sylvan B. Green
Susan G. Groshen
Marek Kimmel
Victor Kipnis
James R. Murphy
Richard A. Olshen
Philip C. Prorok
Timothy R. Rebbeck
Philip S. Rosenberg
Harland N. Sather
Daniel J. Schaid
Mark R. Segal
Richard M. Simon
Donald M. Stablein
Robert E. Tarone
Sholom Wacholder

EDITORIAL ADVISORY BOARD

Hans-Olov Adami
Harry B. Burke
Pelayo Correa
Ethan Dmitrovsky
Merrill J. Egorin
Soldano Ferrone
Isaiah J. Fidler
Curtis C. Harris
Emily L. Harris
Kathy J. Helzlsouer
Maureen M. Henderson
Gloria H. Heppner
Allan Hildesheim
Waun Ki Hong
William J. Hoskins
Peter J. Houghton
David H. Johnson
V. Craig Jordan
Ilan R. Kirsch
Kenneth H. Kraemer
Alexandra M. Levine
Martha S. Linet
Henry T. Lynch
Frank L. Meyskens
Anthony B. Miller
Malcolm S. Mitchell
David L. Page
Kenneth J. Pienta
David Schottenfeld
Richard K. Severson
Jerry W. Shay
Debra T. Silverman
Sandra M. Swain
G. Marie Swanson
Mario Sznol
Raymond Taetle
Ian Tannock
Peter R. Twentyman
Regina G. Ziegler

Richard D. Klausner
Director, National Cancer Institute
Susan Molloy Hubbard
Director, International Cancer Information Center
Julianne Chappell
Chief, Scientific Publications Branch; Executive Editor

EDITORIAL STAFF

Scientific Editors: Barbara Rimer, Dr.P.H.
J. Paul Van Nevel

***Journal* Managing Editor:** Richard E. Manrow

Monograph Coordinator: Elizabeth Horowitz

Manuscript Editors: Elaine Price Beck
Janice Deal
Joan O'Brien Rodriguez
Elizabeth Horowitz

Editorial Assistants: Aylin An, Danielle D. Reed, Patricia A. Smith, Tenisha D. Smith

MARKETING STAFF

Marketing: Joy Cox

Press Contact: Dan Eckstein*

EDITORIAL POLICY: Manuscripts from key conferences dealing with cancer and closely related research fields, or a related group of papers on specific subjects of importance to cancer research, are considered for publication, with the understanding that they have not been published previously and are submitted exclusively to the *Journal of the National Cancer Institute Monographs.* All material submitted for consideration will be subject to review, when appropriate, by at least one outside reviewer and one member of the Editorial Board of the *Journal of the National Cancer Institute.* Opinions expressed by the authors are not necessarily those of the publisher or the editors.

Proposals for monographs should be submitted to the Editor-in-Chief, *Journal of the National Cancer Institute,* Oxford University Press, P.O. Box 31110, Bethesda, MD 20824-1110.

Journal of the National Cancer Institute Monographs are available on request to members of the National Cancer Institute Information Associates Program (one copy of each monograph per member). Monographs are also available from Oxford University Press, Journals Subscription Department, 2001 Evans Road, Cary, NC 27513 USA. Telephone toll-free in the U.S. and Canada, 1-800-852-7323, or 919-677-0977. Fax: 919-677-1714. E-mail: jnlorders@oup-usa.org.

© Oxford University Press

*Contact by telephone at 301-986-1891, ext. 112.

Cancer Risk Communication: What We Know and What We Need to Learn

Sponsored by the National Cancer Institute

December 10–11, 1998

List of Participants

Elaine Bratic Arkin
Porter Novelli
Lisa G. Aspinwall, Ph.D.
University of Maryland
Barbara Bowles Biesecker, M.S.
National Human Genome Research Institute
Jane E. Brody, M.S.
The New York Times
John Burklow
National Cancer Institute
Nelvis Castro
National Cancer Institute
Caron Chess, Ph.D.
Cook College, Rutgers University
Robert T. Croyle
National Cancer Institute
Terry Davis, Ph.D.
Louisiana State University Medical Center
Connie Dresser, RDPH, LN
National Cancer Institute
Ellen Eisner
National Cancer Institute
Baruch Fischhoff, Ph.D.
Carnegie Mellon University
Geoffrey T. Fong, Ph.D.
University of Waterloo
Vicki S. Freimuth, Ph.D.
Centers for Disease Control and Prevention
Meg Gerrard, Ph.D.
Iowa State University
Bernard Glassman, M.A.T.
National Cancer Institute
M. Alfred Haynes, M.D.
Drew Mahoney Consortium
Robert A. Hiatt, M.D., Ph.D.
National Cancer Institute
Elmer E. Huerta, M.D., M.P.H.
Washington Hospital Center
Barnett S. Kramer, M.D., M.P.H.
National Cancer Institute
Matthew W. Kreuter, Ph.D., M.P.H.
Saint Louis University
Elaine A. Leventhal, M.D., Ph.D.
UMDNJ-Robert Wood Johnson Medical School
Howard Leventhal, Ph.D.
Rutgers University
Jeffrey Levine, M.J.
Medcast Networks
Issac Lipkus, Ph.D.
Duke University Medical Center
Anne Lubenow, M.P.H.
National Cancer Institute
Everly Macario, Sc.D.
Porter Novelli
Edward Maibach, Ph.D., M.P.H.
Porter Novelli
Al Marcus, Ph.D.
AMC Cancer Research Center
Kevin D. McCaul, Ph.D.
North Dakota State University
Sherry Mills, M.D.
National Cancer Institute
Nancy J. Nelson
National Cancer Institute
Patricia Newman, M.G.A.
National Cancer Institute
Annette M. O'Connor, RN, Ph.D.
University of Ottawa
Wendy Pachter, Ph.D., J.D.
National Academy of Sciences
Barbara R. Parker
Duke University Medical Center

List of Participants *(continued)*

Maria T. Pavlova, M.D., Ph.D.
U.S. Department of Energy
Barry Portnoy, Ph.D.
National Cancer Institute
Amelie G. Ramirez, Dr.P.H.
Baylor College of Medicine
Scott C. Ratzan, M.D., M.P.A., M.A.
Academy for Educational Development
Jane Reese-Coulbourne, M.S.Ch.E., M.B.A.
National Cancer Institute
Barbara K. Rimer, Dr.P.H.
National Cancer Institute
Joann E. Rodgers, M.S.
The Johns Hopkins Medical Institutions
Amanda Roth, M.S.W., M.P.H.
Porter Novelli
Alexander J. Rothman, Ph.D.
University of Minneapolis
Cristine Russell
The Washington Post
Peter Salovey, Ph.D.
Yale University
Debbie Saslow, Ph.D.
American Cancer Society
Lisa M. Schwartz, M.D., M.S.
Dartmouth Medical School
Betty Spicher
National Cancer Institute
Victor J. Strecher, Ph.D., M.P.H.
University of Michigan
Stephen C. Stuyck
The University of Texas
M. D. Anderson Cancer Center
J. Paul Van Nevel
National Cancer Institute
Sally W. Vernon, Ph.D.
The University of Texas-Houston
Neil D. Weinstein, Ph.D.
Rutgers University
Steven Woloshin, M.D., M.S.
Dartmouth Medical School

Cancer Risk Communication: What We Know and What We Need To Learn — *1999 Number 25*

Contents

Foreword 1
Richard D. Klausner

The Cancer Risk Communication Meeting in Perspective 3
Barbara K. Rimer

Workshop Keynote Address 5
The editors

Introduction of Section: Challenges Inherent in Communicating Cancer Risk Information 5
J. Paul Van Nevel

Why (Cancer) Risk Communication Can Be Hard 7
Baruch Fischhoff

Improving Cancer Risk Communication: a Discussion of Fischhoff 14
Edward Maibach

What Does It Mean to Understand a Risk? Evaluating Risk Comprehension 15
Neil D. Weinstein

Introduction of Section: Overarching Considerations in Risk Communications: Romancing the Message 21
Joann E. Rodgers

Communicating Health Risk to Ethnic Groups: Reaching Hispanics as a Case Study 23
Elmer E. Huerta, Everly Macario

Dealing With Competing and Conflicting Risks in Cancer Communication 27
Matthew W. Kreuter

New Directions for Risk Communication Research: a Discussion With Additional Suggestions 35
Alfred C. Marcus

Introduction of Section: Enabling Informed Decisions About Cancer Risk 43
Barbara Bowles Biesecker

Treating People With Information: an Analysis and Review of Approaches to Communicating Health Risk Information 44
Alexander J. Rothman, Marc T. Kiviniemi

Cancer Screening Decisions 52
Kevin D. McCaul, Heather E. Tulloch

Risk Communication in Genetic Testing for Cancer Susceptibility 59
Robert T. Croyle, Caryn Lerman

Decision Aids for Patients Considering Options Affecting Cancer Outcomes: Evidence of Efficacy and Policy Implications 67
Annette M. O'Connor, Valerie Fiset, Catherine DeGrasse, Ian D. Graham, William Evans, Dawn Stacey, Andreas Laupacis, Peter Tugwell

Population Risk, Actual Risk, Perceived Risk, and Cancer Control: a Discussion 81
Howard Leventhal, Kim Kelly, Elaine A. Leventhal

Matching Strength of Message to Strength of Evidence: a Discussion 85
Barnett S. Kramer

LIBRARY
OCT 12 1999
National Institutes of Health

(Contents continued on next page)

Introduction of Section:
Persuasion for the Purpose of Cancer Risk Reduction: Understanding Responses to Risk Communications 88
Lisa G. Aspinwall

The Effect of Risk Communication on Risk Perceptions: the Significance of Individual Differences 94
Meg Gerrard, Frederick X. Gibbons, Monica Reis-Bergan

Risk Perception and Risk Communication for Cancer Screening Behaviors: a Review 101
Sally W. Vernon

Persuasion for the Purpose of Cancer Risk Reduction: a Discussion 119
Peter Salovey, Tamera R. Schneider, Anne Marie Apanovitch

Introduction of Section:
Implications for Improving Risk Communication Through Various Channels 123
J. Paul Van Nevel

Risk Communication in Clinical Practice: Putting Cancer in Context 124
Lisa M. Schwartz, Steven Woloshin, H. Gilbert Welch

Interactive Multimedia and Risk Communication 134
Victor J. Strecher, Todd Greenwood, Catharine Wang, Dana Dumont

Is There a Use for Tailored Print Communications in Cancer Risk Communication? 140
Barbara K. Rimer, Bernard Glassman

The Visual Communication of Risk 149
Isaac M. Lipkus, J. G. Hollands

Risky Business—Communicating Scientific Findings to the Public 163
Jeffrey Levine

Living Can Be Hazardous to Your Health: How the News Media Cover Cancer Risks 167
Cristine Russell

Communicating Cancer Risk in Print Journalism 170
Jane E. Brody

Challenges to Improving Health Risk Communication in the 21st Century: a Discussion 173
Geoffrey T. Fong, Lynn A. Rempel, Peter A. Hall

Implications for Improving Risk Communication Through Various Channels: a Discussion 177
Cristine Russell

Introduction of Section:
Breakout Session Reports 179
J. Paul Van Nevel

Cancer Risk Communication—What We Need to Learn 179
Edward Maibach

Cancer Risk Communication—What We Know 182
Elaine B. Arkin

Foreword

Richard D. Klausner

Over the past decade, our understanding of cancer genetics has been transformed, and we have gained remarkable new insights into the causes of cancer. Almost every scientific advance results, at some point, in a public discussion about cancer risks, so it is sobering to realize that we probably know less now about how to communicate cancer risks than we know about cancer genetics. Moreover, decisions by both individuals and health professionals about cancer prevention, detection, and treatment should be an informed balancing of potential results with potential risks. And, as scientists, we have a much clearer picture of individual risks than ever. But here is the paradox: We know far less about how to communicate risks than we know about how to calculate them.

Because of the importance of this topic, and because so little is known about how to communicate cancer risks effectively, I strongly supported the National Cancer Institute's sponsorship of an important conference on "Cancer Risk Communication: What We Know and What We Need to Learn." Publication of the papers presented at that conference as a monograph of the *Journal of the National Cancer Institute* adds important substance to the knowledge base on cancer risk communication. This monograph has a broad array of contributors, including renowned scholars whose field of study is risk communication and those developing models for use by health professionals. The monograph also includes contributions from representatives of special populations, medical journals that publish risk papers, the mass media, and institutions that conduct risk-related research and then communicate results of that research. As a result, the monograph not only is a major contribution to the literature about cancer risk communication, it provides a unique compilation of expertise.

Just as important, the monograph will help guide future research because it identifies gaps in knowledge that must be filled. Until that can be done, we must rely on the best practices now known about how to communicate cancer risk. These practices are identified here and can be adopted immediately by those who disseminate cancer risk information.

Questions about how to communicate cancer risk are plentiful, but they are only a fraction of the questions about cancer communications that confront us today. How can we promote the demand for access to and use of cancer information given the high national rates of medical and scientific illiteracy? How can we assure that cancer communications are salient, relevant, and culturally sensitive to diverse audiences? How can we better design our interventions to learn what works, what does not work, and why? How can we help physicians, who have limited time to spend with patients, maximize their communication about cancer? How can we redesign our information systems so they give people the information they want, how they want it, and when and where they want it?

Now, thanks to emerging communications technologies and knowledge developed through academic research, we have an extraordinary opportunity to answer such questions, and we must seize it. Each year, the National Cancer Institute prepares a budget proposal that identifies extraordinary opportunities to further progress brought about by earlier research successes. Our proposal for 2001 will include cancer communications as an extraordinary opportunity for investment. Such an investment can shape the emerging national information infrastructure to improve cancer communications and thereby help improve the public's health. This monograph is an important step in that direction.

Affiliation of author: Director, National Cancer Institute, Bethesda, MD.
Correspondence to: R. D. Klausner, M.D., National Institutes of Health, Bldg. 31, Rm. 11A48, MSC 2590, Bethesda, MD 20892.

The Cancer Risk Communication Meeting in Perspective

Barbara K. Rimer

The National Research Council *(1)* defined risk communication as an interactive process of exchange of information and opinion among individuals, groups, and institutions. It involves multiple messages about the nature of risk and other messages, not strictly about risk, that express concerns, opinions, or reactions to risk messages or to legal and institutional arrangements for risk management. When applied to cancer, this is a useful working definition of risk communication.

Cancer risk communication is a vital topic for many groups, including the media, health care professionals, researchers, and, of course, the public. Media professionals are faced daily with the need to communicate risk. Nowhere were the challenges and repercussions of cancer risk communication more apparent than when we have attempted to communicate the lifetime risk for breast cancer (the well-known one-in-eight number). As Bunker et al. *(2)* recently stated, people do not understand the idea of lifetime risk. They also do not understand that the average risk is modified by a number of individual factors, such as age and family history of cancer. This points out one of the important problems in communicating cancer risk information. People have trouble understanding numbers, especially when the numbers are probabilities.

But perhaps even harder is communicating the concept of eompeting risks. For example, women focus so much on breast cancer risks that they underestimate the risks of heart disease and lung cancer. Risk communication may be made harder by the influence of certain advertisers (e.g., tobacco companies). A survey by the American Council on Science and Health *(3)* showed that the risks of smoking tend to be downplayed in women's magazines. The survey found that articles on tobacco represented fewer than 1% of all health-related articles. No wonder people have problems perceiving tobacco-related risks correctly. Recently, Nelkin *(4)* argued, "The media response to research reports about health risks suggests that efforts to reach the public must consider the important differences between lay and professional understanding of certain concepts. To scientists, evidence is a statistical concept, and the focus is on aggregate data. But to laypersons—and many journalists—anecdotal information and individual cases are also compelling evidence. The media focus is on the immediate concerns of readers. Will I be harmed?" Balancing these different perspectives about scientific evidence is an important challenge of cancer risk communication.

Researchers are grappling with a number of important questions related to cancer risk communication, including the following: What is the role of risk perceptions in cancer-related behaviors? What effects do risk communications have on decision making and behavior? How do we communicate risk in numbers, words, and graphics?

Practitioners are faced with patients who may want to know their own risk of breast cancer and what they can do about it. Other patients (e.g., those who smoke) may want to avoid discussion of risk. And, of course, the possibility of testing for genetic susceptibility to cancer raises many other patient questions about risk.

This meeting was designed with the ambitious goal of trying to bridge the communication perspectives of researchers, the media, and practitioners. We planned to emerge with a clearer understanding of what we know about cancer risk communication and what we need to know to bridge what often seems like a chasm between research and practice. That is why we asked the participants to focus on the two questions: 1) What do we know? 2) What do we need to know about cancer risk communication?

The importance of this topic was highlighted in a recent report by the Institute of Medicine and the National Research Council *(5)* that dealt with recommendations for people exposed to iodine-131. The committee suggested that the U.S. Department of Health and Human Services (DHHS) consider funding research to develop a better understanding of how people perceive the benefits and harms of cancer screening and how those perceptions are affected by different ways of presenting quantitative information and different ways of structuring patient–clinician communication. They emphasized that DHHS would be well served by efforts to understand better how individuals make judgments about risk situations and, more specifically, how they use information to do so. This is a useful statement of research needs for the broader field of cancer risk communication as well as the specific case of iodine-131.

This meeting of some of the United States' leading researchers and communicators represents an important opportunity to advance both the research and the practice agendas related to cancer risk communication. Our goals, after all, are more similar than different—to improve the communication and comprehension of cancer risk.

References

(1) National Research Council. Improving risk communication. Washington (DC): National Academy Press; 1989.

(2) Bunker JP, Houghton J, Baum M. Putting the risk of breast cancer in perspective. BMJ 1998;317:1307–9.

(3) You've come a long way . . . or have you? Popular women's magazines and smoking—still downplaying risks. New York (NY): American Council on Science and Health; 1998.

(4) Nelkin D. Scientific journals and public disputes. Lancet 1998; SII25–8.

(5) Institute of Medicine and National Research Council. Exposure of the American people to iodine-131 from Nevada nuclear-bomb test: review of the National Cancer Institute report and public health implications. Washington (DC): National Academy Press; 1998.

Affiliation of author: Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.

Correspondence to: Barbara K. Rimer, Dr.P.H., National Institutes of Health, Executive Plaza North, Rm. 242, Bethesda, MD 20892 (e-mail: brimer@nih.gov).

Workshop Keynote Address

The editors

The workshop's keynote address, by Scott C. Ratzan, M.D., Editor-in-Chief of *Health Communication Technology & Educational Innovation,* was published earlier in another periodical, so it is not included here; however, the Monograph editors are pleased to provide the reference to this important paper:

Ratzan SC. Cancer risk communication: what we know and what we need to learn. Vital Speeches of the Day 1998;65:267–70.

Introduction of Section: Challenges Inherent in Communicating Cancer Risk Information

J. Paul Van Nevel

In this group of papers, Fischhoff and Weinstein, with a response by Maibach, examine the risk communication challenge from the perspective of how people process information. The authors focus on the cognitive barriers to understanding cancer risks. These papers make the case that successful risk communication depends on the ability of those who originate risk communications to understand and address audience members' perceptions of and feelings about the risk in question. The authors postulate that successful communication of risk information poses challenges not only for communications planning but also for creating links between the public and the science underlying risk information. They also present new directions in risk communications research.

Affiliation of author: Office of Cancer Communications, National Cancer Institute, Bethesda, MD.

Correspondence to: J. P. Van Nevel, National Institutes of Health, Bldg. 31, Rm. 10A31, MSC 2580, Bethesda, MD 20892 (e-mail: nevelp@occ.nci.nih.gov).

Why (Cancer) Risk Communication Can Be Hard

Baruch Fischhoff

Effective risk communication uses audience members' time well by providing them with the information that they most need, in a form that they can easily comprehend. Accomplishing this task can be hard because of problems with both the transmitter and the receiver. The former must determine what is most worth saying. The latter must integrate that message with their often fragmentary mental models of the processes creating and controlling the risks. One strategy for improving communication is to use analytic methods for selecting the information to transmit, based on its criticality to recipients' decision making. A second strategy for improving communications is adapting the message to the cognitive processes of its recipients. Together, these strategies can reveal the limits to communication and how best to work within them. [Monogr Natl Cancer Inst 1999;25:7–13]

Understanding most risks has a qualitative and a quantitative component. Quantitatively, those exposed to a hazard need to know how big the threat is, in terms that allow them to evaluate its seriousness (compared with the costs of reducing it). Qualitatively, they need to know how the hazard can be controlled, in terms that allow them to select and implement appropriate actions as well as to detect changes in their circumstances. These two sets of information needs are interrelated. The magnitude of a risk should affect people's interest in its mechanisms. Conversely, an understanding of those mechanisms may make quantitative estimates more meaningful, by showing how the risk could be that big or small.

Like any good communication process, formulating an effective risk message begins by characterizing the information needs of its intended audience. A message should focus on the most critical facts that have yet to be understood. Then those facts need to be transmitted in a credible, comprehensible way. Accomplishing these goals requires a *normative* analysis of which facts are most worth knowing and a *descriptive* analysis of what people currently believe. The normative analysis should apply the best scientific knowledge about risk to the specific concerns of the target individuals. The descriptive analysis should apply the best scientific knowledge about human information processing to those individuals' specific needs. The resulting communication should be tested and the process continued, until audience members experience no more than the acceptable level of misunderstanding *(1)*.

The next two sections of this paper discuss, in turn, communicating qualitative and quantitative information. Each process requires (*a*) normative analyses of what is worth knowing and (*b*) descriptive studies of what is worth saying. The concluding section considers the strategy of relying on informational messages to help people cope with serious risks.

What Is Worth Knowing

Quantitative Information About Risks

Conceptually, determining quantitative information needs is straightforward.

1) Describe the decisions that audience members face. The conventional way to do that is with a decision tree, like that in Fig. 1. It begins with the options available and ends with the consequences whose probability of occurrence is affected by the choice of option. Between them lie uncertain events, whose outcome determines whether the consequences do, in fact, occur *(2,3)*.

2) Summarize the target individuals' current perceptions of these decisions. Conventionally, these beliefs are represented as the subjective probabilities of the possible outcomes (e.g., the chances of true and false positives from a mammogram or of living various time periods with a chosen treatment). Individuals' values are represented by the perceived utility (or disutility) of experiencing the different consequences (or, equivalently, the acceptable trade-offs among them).

3) Identify the optimal choice implied by these judgments. This could be called the subjectively rational choice, in the sense of representing what people would choose if they made best use of their (possibly misinformed) beliefs.

4) Redo the analysis, replacing each subjective estimate, in turn, by the best available scientific estimate. For questions of fact, this step means changing the probabilities to ones informed by current science. For questions of value, this step means changing the utilities to ones informed by a full appreciation of what it means to experience the consequence (which might, mercifully, be unprecedented in many people's lives).

Replacing all model estimates with the best scientific estimates produces what could be called the objectively rational choice. (This use of "objective" recognizes that science always involves an element of judgment.) Those estimates whose replacement has the largest effect on choices are the logical foci of communication (i.e., they are the facts whose transmission matters most) *(4)*. In Fig. 1, that choice might mean giving a more accurate feeling for the effectiveness of chemotherapy or the associated nausea. Or it might mean pointing to options or consequences that were neglected entirely (e.g., a new clinical trial, career changes).

Message content should proceed as far down the list of priorities as the communication opportunity allows. Inevitably, there are limits. Recipients may lack the time or cognitive capacity to process every fact that might help them—or the emotional capacity to deal with some specific facts. The communicator may not feel free to discuss some issues (e.g., effects on sexual performance, risks of professional misconduct). Thus, effective communication must be realistic about where people are coming from and about how far they can be taken.

Getting information out of people, regarding their current beliefs, involves much the same cognitive processes as getting

Correspondence to: Baruch Fischhoff, Ph.D., Department of Social and Decision Sciences, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213 (e-mail: baruch@cmu.edu).

© Oxford University Press

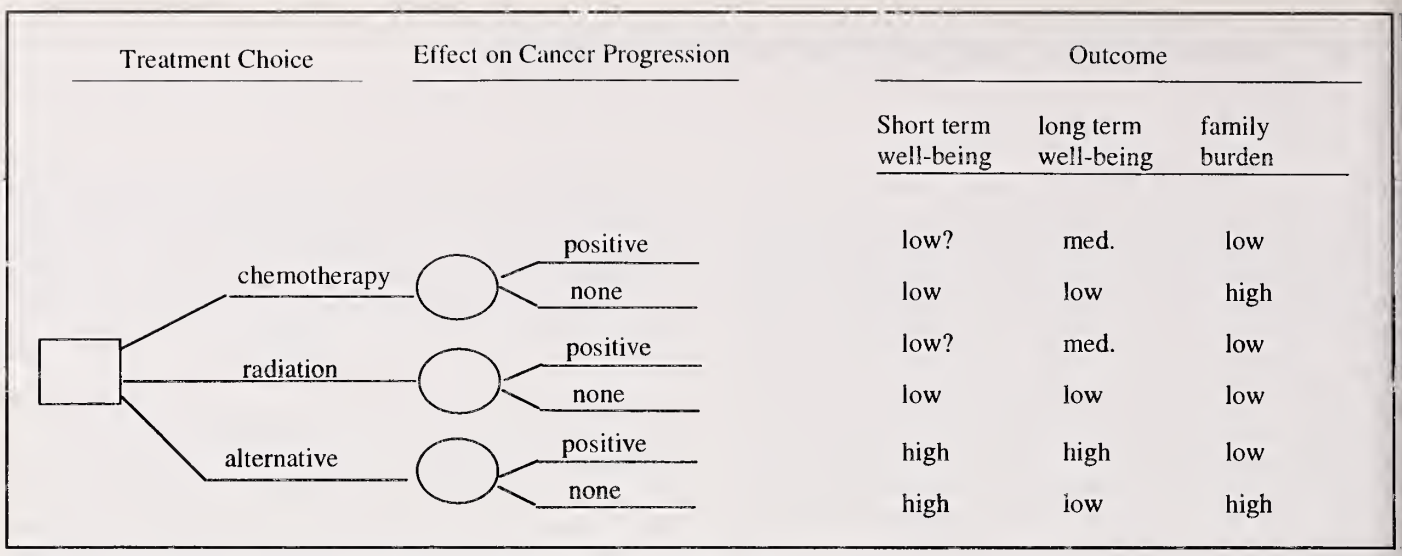


Fig. 1. A partial decision tree for cancer treatment. Actual values depend on specific case.

information into them. As a result, it is a major topic of this paper. Although eliciting expert beliefs is left to other sources *(5,6)*, it faces analogous cognitive issues. Identifying and integrating the full set of expertise relevant to an analysis often poses significant intellectual (and institutional) challenges. Nonetheless, that discipline is needed before communicators can responsibly demand people's attention, occupy the spot in their lives that might be filled by other (better) messages, or despair of their incompetence (for failing to learn from messages that were, in truth, not very good).

Qualitative Information About Risk

Identifying qualitative information needs follows from identifying quantitative information needs. An influence diagram *(7,8)*, like that in Fig. 2, is created to show the determinants of each uncertain event in the decision tree. In it, each node represents a variable; an arrow means that the variable at the head depends on the one at the tail. Purely qualitative models "merely" assemble the relevant risk factors and the processes linking them—in a sufficiently rigorous way to allow running

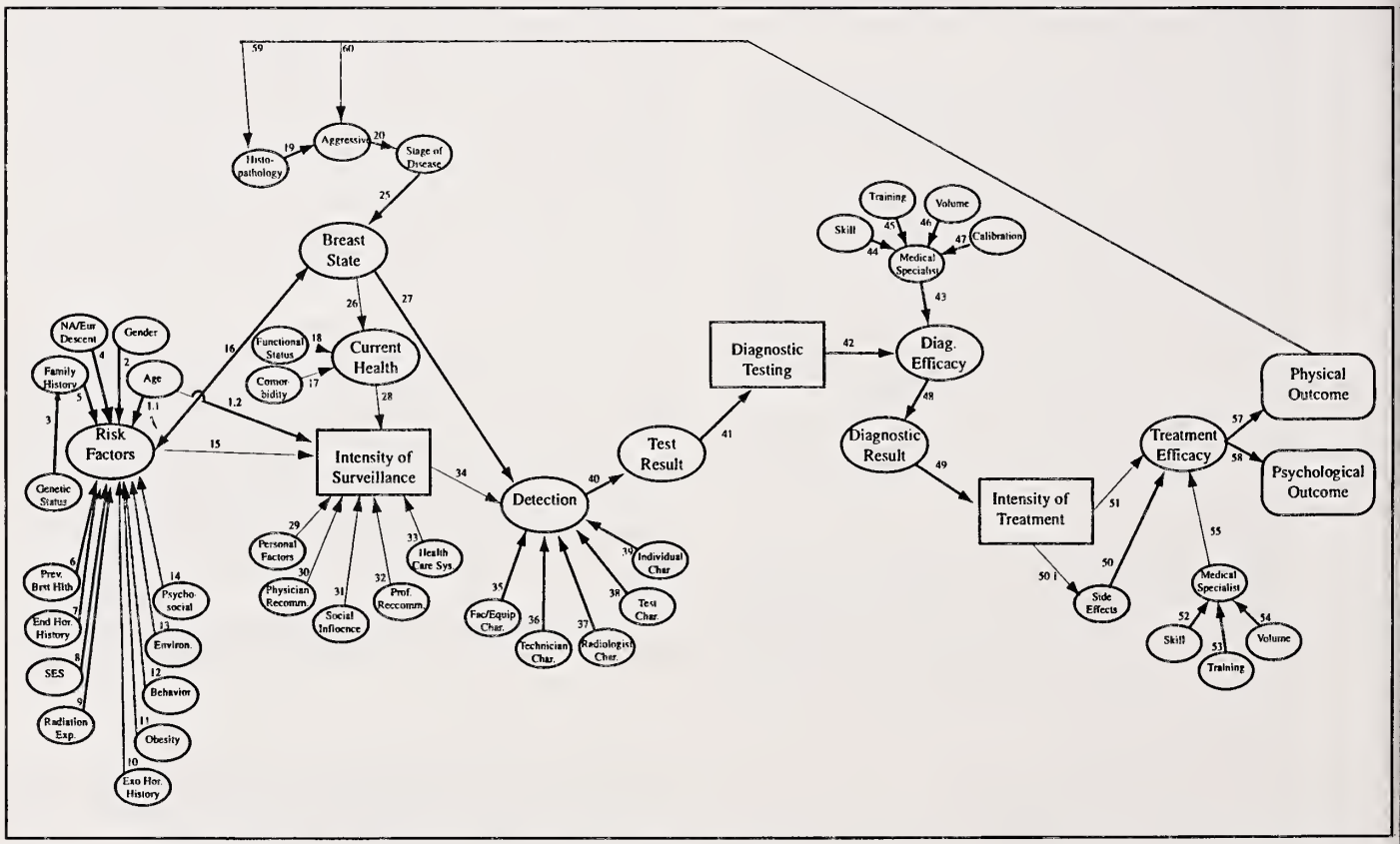


Fig. 2. An expert model for breast cancer diagnosis and treatment. Used with permission *(45)*.

the numbers at some later day. Nonetheless, even the most qualitative modeling involves some rough quantification, to screen factors with weak influences or dubious scientific support.

Setting communication priorities follows the same logic as with quantitative information: (*a*) create influence diagrams relevant to the pending choices of target audience members; (*b*) conduct descriptive studies, identifying current beliefs about relationships in the diagram; (*c*) assess the risk judgments following from those beliefs; (*d*) identify the missing facts that would most improve those judgments; and (*e*) focus communications on them.

Those communications might involve correcting erroneous beliefs, perhaps by describing the strength of a relationship (e.g., "diet has only been weakly linked to breast cancer"), perhaps by explaining its rationale (e.g., "the body has complex metabolic processes, mediated by genes, activity, and environment, making it hard to tell what difference diet makes, especially with health problems whose natural progression is poorly understood"). Those communications might involve pointing to currently ignored relationships or refuting mistakenly accepted ones. Or they could provide background information, clarifying the picture as a whole (e.g., how hormones or breast mass change with age).

We call the result of a normative analysis an expert model and the result of descriptive research a mental model *(1,9)*. In these terms, risk communication hopes to make lay people more like experts, in those ways that matter most to lay decisions. Because these terms may connote more than their literal intent, a few words of terminologic clarification are in order.

Model suggests a formal representation onto which one can map objects and operations. The expert model should allow such mapping; lay mental models often will not, especially for complex, novel problems (like many risks).

Expert refers to those who know most about each relationship in the model, without implying that any one expert is the best source for all topics nor that the community of experts knows everything.

Lay denotes the source of audience members' knowledge, not its extent. The experiences, curiosities, and distractions of everyday life can leave lay people clueless, self-educated, radically skeptical, or shrewdly observant. Often, lay people's indigenous technical knowledge *(10)* is essential for connecting general messages to specific circumstances.

Quantitative Lay Knowledge

Assessment

Assessing beliefs about uncertain contingencies means eliciting numeric judgments for the probabilities of clearly defined events. One common source of ambiguity is verbal quantifiers, such as "likely" and "rarely," whose numeric equivalent can vary by context and by individual (e.g., "likely" to rain versus "likely" to have a painful side effect) *(11,12)*. Fortunately, eliminating it is straightforward: offer a linear scale, anchored at "0% = can't happen" and "100% = certain to happen," and label the appropriate tick marks with 10, 20, . . ., 90). If the probabilities might be very small, offer a log expansion of the section between 0% and 1%, as in Fig. 3. Such judgments have been elicited for many events, with diverse (and occasionally representative) populations of adolescents and adults. They require a little more work from respondents than do verbal quantifiers. As a result, they are less popular as a way to provide one's own beliefs but more popular as a way to receive others' beliefs *(13)*. Their increased interpretability should justify a little respondent discomfort.

Two cautions are in order when probabilities are elicited with open-ended questions, as might occur with phone interviews. (*a*) People sometimes say "50," when they mean "fifty-fifty," indicating an inability or unwillingness to answer, rather than a numeric value *(14)*. This tendency may be relatively common with poorly understood, negative events (the topics of many risk surveys). Fifty-fifty responses convey information about respondents' internal states; however, they exaggerate the perceived probability of unlikely events, if they are interpreted as numbers. (*b*) People's preference for integer responses also affects estimations of very low (and very high) probabilities. Such floor (or ceiling) effects are special cases of familiar psychophysical phenomena *(15)*.

Resolving ambiguity in event descriptions is more complicated, insofar as there are many more words than kinds of numbers. One way to address it is the following process:

1) Ask people to think aloud as they answer a draft version of

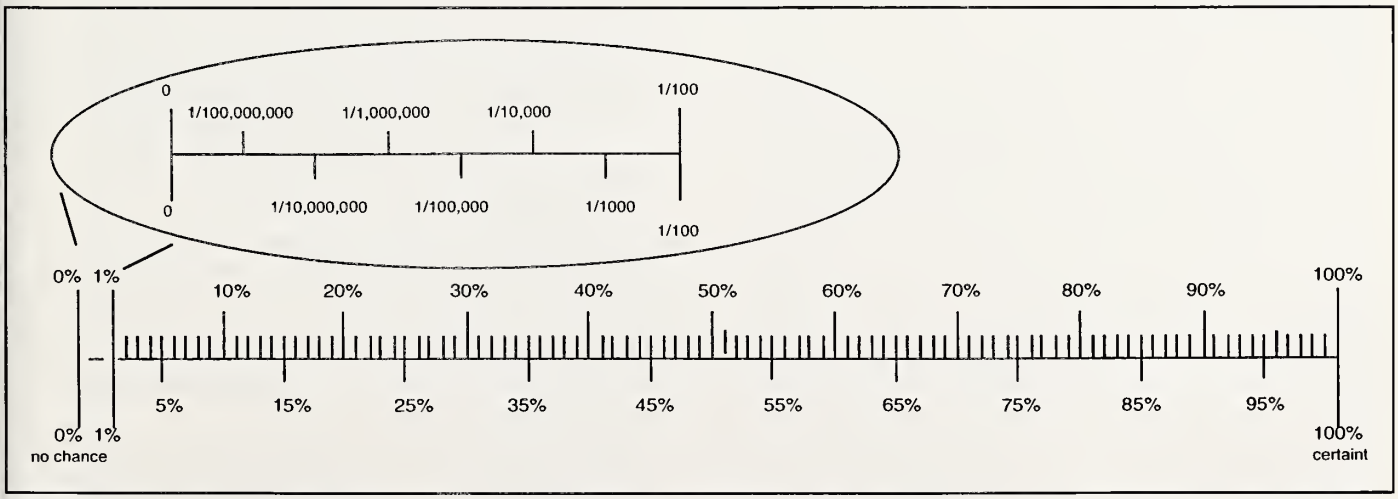


Fig. 3. A log-linear scale for eliciting probabilities. For the sake of balance, one might also expand the 99%–100% range. One variant uses a magnifying glass to highlight the very small probabilities *(46)*.

the question (e.g., "What is the probability of getting chlamydia from having sex"?). Note each uncertainty raised or assumption made about its definition (e.g., "Do you mean the risk of doing it once or many times?" "Well, it's different for boys and girls").

2) Ask respondents to elaborate on each such concern to clarify its meaning and prompt other issues.

3) Code responses in terms of the risk factors that they reflect. Table 1 offers one general coding scheme for risk behaviors *(16,17)*.

4) Draft a test question clearly specifying the needed details. Include factors relevant to lay people even if irrelevant to technical experts. If a question ignores an issue that respondents consider important, they may impute a value to it anyway.

5) Conduct tests to ensure that the fully specified question is understood as intended. That can be done with (*a*) think-aloud protocols (repeating steps 1 and 2), (*b*) manipulation checks (asking respondents to recall task details or make simple inferences from them), and (*c*) confirmatory tests, having subjects answer versions of the question that differ in each risk factor that was important to think-aloud subjects (to see if people really attend to the details that they say matter to them).

The final choice of wording may force some hard choices. For example, we might want to ask very general questions to cover a large set of events. However, doing so poses a quandary for respondents who make finer distinctions. To answer, they must either focus on some subset of the people and situations covered by the question (as defined by categories like those in Table 1) or compute a weighted average in their heads. Unlike probability scales, which can be developed once and then applied to many events, this procedure must be repeated for each event. Fortunately, the research on any term can then be applied whenever it occurs (as long as the target populations have similar linguistic norms). Thus, for example, McIntyre and West *(18)* have studied lay interpretations of safe sex, Turner et al. *(19)* have done so for drug-related behaviors, Leigh *(20)* has examined alcohol-related terms, and Morwitz *(21)* has reviewed verbs related to behavioral intentions (e.g., "will purchase"). A substantial amount of literature considers the meanings of "risk" itself *(22–24)*.

Table 1. Coding framework for drinking and driving question

Framework element	Risk factor categories	Example variables
Behavior	Dose	Amount of alcohol consumed
	Amount	
	Potency	
	Method	
Other behaviors		
	Risk buffers	Amount of food eaten
	Risk amplifiers	Other drugs consumed
	Time-related	Night or day; day of the week
	Place-related	Where alcohol was consumed
Actor	Physical	Tolerance to alcohol
	Cognitive	Awareness of effects of alcohol
	Social–psychological	Mood
	Material	Wealth
	Spiritual	Faith
	Skill	Driving skill
	Character	Responsible; mature
	Age	
	Sex	
	Genetic history	
	Status	
	Luck	
	Motivation	
	Other	
Context	Social	
	General, cultural	Drinking norms
	Family	Family approval
	Peers, others	Peer approval
	Environmental	Road conditions
Risk outcome	Social reactions	Get in trouble
	Personal effects	
	Physical	Injury
	Psychologic	Worry, guilt
	Cognitive–physiological	Kill brain cells
	Cognitive–psychological	Can't think
	Material	Lose car, lose license
	Accidents	Get in a wreck while high
	Lifestyle	Become a bum
	Complex	Get high
	Effects on others	Hurt your friends, family
	Behaviors	Use more, do heavier drugs
	Severity, type when measured	

Improvement

The cognitive processes involved in expressing beliefs are also involved in creating them. Thus, similar design principles apply here. For example, communications should satisfy a person's need and preference for numeric quantifiers. Like response modes, information displays send signals. Thus, a log-linear display connotes that very small probabilities are possible and that one should take the trouble to distinguish among them. Its appropriateness is a matter of context and empirical testing.

Recommending numeric displays may seem odd, given the frequent lamentations over lay innumeracy *(25,26)*. Although a full analysis is impossible here, one should note that studies showing lay trouble with numbers rarely speak to the present issue. These studies typically examine how people combine information from diverse sources, not how they understand individual facts. They focus on what people currently know, not on what they can learn. Furthermore, claims of lay incompetence frequently rest on untested assumptions (e.g., nonexpert risk judgments reflect differences in beliefs rather than in definitions of "risk"). Expert ambiguity can cause behavior that appears like innumeracy. For example, in the late 1970s, anecdotal evidence suggested public discomfort with probability-of-precipitation (PoP) forecasts; the problem, however, proved to be confusion over what forecasters meant by "precipitation," not "probability" *(27)*. Finally, studies *(28–30)* showing lay innumeracy typically use highly structured tasks; whether providing the contextual cues of everyday life helps or hinders particular judgments is an empirical question.

Whatever the difficulties, communicating quantitative information is a matter of necessity. Effective decision making requires quantitative knowledge. We need to deliver it comprehensibly, so people need not guess about the size and probability of effects. We can hope that repeated exposure to standard displays of risk measures creates public understanding, as it has with other numbers now part of everyday life (e.g., PoP, recommended daily allowance).

Whatever quantifiers are used, the associated events must be clear. One possible strategy is using a controlled vocabulary. A push in this direction may come from the U.S. Food and Drug Administration's proposed warning label for over-the-counter

drugs *(31)*. Although it avoids numbers (and has little empirical validation), this proposal could promote a common terminology for this large class of risks. An analogous (but quantitative) proposal has been advanced for automobile safety *(32)*. All proposals need testing to confirm designers' intuitions.

Clear, consensual definitions may, however, not suffice if people cannot understand that kind of information. Such difficulties are the stock in trade of behavioral decision making *(28,33)*, which studies lay understanding of uncertain processes that are neither self-evident nor routinely taught. Three examples may give a flavor for these problems and their possible treatment.

1) The acceptability of very low probability risks can depend on how they are presented (e.g., relative versus absolute risk, frequency per year versus years between occurrences). The fairest presentation may be to show multiple perspectives, allowing recipients to triangulate and seek a rounded feeling for just how big the risk really is *(34,35)*.

2) People have difficulty translating between single and multiple exposures to a hazard. Typically, they underestimate how quickly small risks (and small risk reductions) add up. Doing the math for them may be the only way to ensure that they get both perspectives (even if they cannot quite reconcile the two in their minds) *(36,37)*.

3) People tend to anchor their beliefs on initial estimates *(28)*. One protection is to present ranges rather than best guesses. The credibility of those ranges might be increased by explaining why the science is that uncertain (e.g., poor data, complex processes, competing claims). Such disclosure can prepare lay people for hearing competing arguments and seeing discrepant anecdotal evidence.

Qualitative Lay Knowledge

Assessment

Determining qualitative information needs begins by comparing lay and expert beliefs, regarding the factors creating and controlling risks. The sources of discrepancies can vary. Lay mental models may assign central roles to factors that technical experts have either rejected on the basis of evidence or neglected as a result of disinterest. Lay people may have insights that are lacking in expert models, drawing on a limited set of disciplines. For example, a modeling team might lack specialists in nutrition, social support, psychoneuroimmunology, or alternative medicine—all factors that might play roles in lay thinking, if not in the actual course of disease. For example, women's mental models of the recent controversy over mammography screening for women in their 40s seem focused more on their perceptions of health care finance than on scientific uncertainties *(38)*.

Our approach to determining the elements of lay mental models parallels that for determining lay definitions of terms in quantitative estimates: (*a*) Ask individuals to, "Tell me what you have heard about *X*. Don't worry whether it is right or wrong. I'm just interested in what is on people's minds. The goal of our project is to help them to understand things better." (*b*) Ask them to elaborate on everything they say, and then elaborate on those elaborations. (*c*) After exhausting the topics that arise spontaneously, ask about domains in the expert model that should be common knowledge but have yet to be raised (e.g., "You haven't said anything about treatment, what can you tell me?"). (*d*) Code the interview in terms of the expert model, augmented, as needed, with nonexpert concepts. (*e*) Conduct confirmatory tests to validate these mental models [e.g., structured surveys, problem-solving exercises *(1,9)*].

The beliefs that respondents express can be characterized in terms of prevalence (which issues occupy people), accuracy (where do they have expertise), calibration (how appropriate is their confidence in their beliefs), co-occurrence (which ensembles of beliefs must be addressed together), and content (what is the substance of those beliefs).

Improvement

The communications following from these analyses address those mental model elements most critical to the decisions facing audience members. These messages may present the entire expert model, perhaps with minimal detail (when audiences know nothing), perhaps with great detail (when audiences know a lot but cannot keep the pieces straight). Or messages may focus on specific facts, perhaps addressing "bugs" that undermine otherwise solid mental models, perhaps providing a stopgap for those knowing little. Four examples from our own work may illustrate these possibilities:

1) The carcinogenic potential of electromagnetic fields is a recurrent public concern. Our descriptive research found that most citizens knew so little that they needed basic orientation, whereas a few had much, albeit, somewhat fragmentary knowledge. Our brochure was tiered, so that it could be read at several levels of detail (for readers with varying expertise and interests). It made a special effort to debias common lay misconceptions regarding how quickly electromagnetic fields decline with distance and how they can be shielded. We tried to clarify public controversies by explaining why scientists disagree about electromagnetic fields. Members of each camp reviewed how we represented their positions. We have sold (at cost) a couple hundred thousand copies of this brochure and would like to think that it has helped to focus the national debate *(39)*.

2) Space missions far from the sun or requiring concentrated power currently must use nuclear energy sources. After seeing citizen opposition to the Galileo shot (to Jupiter and beyond), we conducted a series of descriptive studies, drawing most participants from a local conservation group. They were found to have (*a*) better elaborated mental models for the effects of potential problems than for their sources, (*b*) a tendency to confuse radiation and radioactivity, and (*c*) coherent beliefs (and preferences) regarding alternative designs for response to orbital failure. The research led to a brochure that left test subjects better informed about and more favorable toward the technology than did the National Aeronautics and Space Administration's (NASA) brochures *(40,41)*. Our research did not seem to influence NASA's communications, although the agency did refer some journalists to us when the Cassini shot generated similar controversy.

3) In the late 1980s, the U.S. Environmental Protection Agency (EPA) began an ambitious campaign to reduce domestic radon exposures, which were thought to be a major source of lung cancer (especially in homes with smokers). In an unusual (and laudatory) move, EPA evaluated its intervention. Unfortunately, little increase in testing or remediation was done. Our descriptive studies *(9)* suggested that, although much of EPA's message had gotten through, citizens' mental models often had a major bug: Reasonably but erroneously, they believed that radon meant permanent contamination. If so, then there was

little point in testing. The revised EPA *Citizen's Guide to Radon* addressed this concern. In this case, the relevant science is easy to explain: Radon decays very rapidly. As a result, a small amount can damage your lungs. However, it is also gone quickly. So the problem vanishes if homeowners can keep additional radon from coming into their houses.

4) Methylene chloride, used as paint stripper, is one of many suspected carcinogens used in the home. Consumers' exposure depends on their usage patterns. Unfortunately, the package labels that might shape their behavior lack clear wording and organization. By use of a combination of formal modeling and mental models interviews, we identified supply curves for usage instructions, prioritizing message content. The model allows different assumptions regarding how many instructions users will read and how full their compliance will be—given their mental models and perceived need to get the job done *(42,43)*.

Each such application must consider the particulars of that cancer risk. Yet, each can take advantage of common components: the science (e.g., indoor air circulation, examples 2 and 4), the perceptions (e.g., confusing radiation and radioactivity, examples 2 and 3), or the medium (e.g., brochure design, examples 1, 2, and 3). As a result, one challenge to the managers of communication programs is creating opportunities for such modular design. My colleagues and I hoped to advance this goal in the area of breast cancer by creating the overall expert model of Fig. 2 and by conducting mental model interviews on the many topics in its domain. The voluminous literature on science education provides advice on conveying many specific processes (e.g., hormones, doubling rates). The expert model is intended to facilitate mapping that research into communication-relevant terms.

Conclusion

Cancer risk communication can be difficult because people are imperfect. Their knowledge of the relevant science is often fragmented and inaccurate. Neither their education nor their everyday experience ensures the numeracy or scientific literacy demanded by these complex, dynamic, and uncertain processes. Even when they have the cognitive resources, people may care too little to think at all about distant risks; they may care too much to think clearly about imminent ones.

In other cases, though, when communication fails, the difficulty does not lie with the receivers. Rather, the transmitters have failed to understand their audience. The transmitters repeat what the receivers know and neglect what they do not. Transmitters patch minor holes and neglect major ones. They use jargon or loaded language. They speak before they listen and undermine their credibility.

Informing the public should follow the same conversational norms as informing children, patients, or students: Figure out which facts have the greatest value to the audience; then relay those facts clearly. Sometimes this step will mean a brief statement; sometimes it will require providing needed context. Explicit analysis and empirical testing are needed, when we are too distant from our audience to get messages right the first time or to evaluate our success. The present approach uses decision-theory formalisms to create a platform capable of integrating results from diverse risk sciences (into the expert model); it uses behavioral science techniques to characterize lay beliefs (as mental models); it draws on research in judgment and science education to address critical gaps.

Creating a computable expert model can impose valuable discipline, even when no numbers are run, by forcing the scientists involved to think through the relationships among their fields. It can also help scientists play more constructive roles in the communication process. Creating a computable expert model means listening to the scientists first—always a good step in communication. It decomposes the communication task, so that it seems more feasible (especially for scientists who would otherwise have but a blurred image of a hysterical public). The computable expert model also puts scientists' personal interests in a broader setting, providing a way to coordinate risk research programs and focus them on communication (i.e., public) needs *(44)*.

Given the emotions surrounding cancer risks, greater formalization may seem like a strange strategy. Clearly, cancer risk communication is not just about facts. The interpretation of a message may cause and be affected by powerful emotions. The hope here is to get the information part right. Doing so allows people to be as cognitive as they want to be in their decision making. It may encourage them to think more before they act, by making thought seem more tractable. Just as undue emotion can swamp cognition, ineffective cognition can generate emotion—as people despair of thinking their way through to answering their problems.

References

(1) Fischhoff B, Bostrom A, Quadrel MJ. Risk perception and communication. In: Detels R, McEwen J, Omenn G, editors. Oxford textbook of public health. London (U.K.): Oxford University Press; 1997. p. 987–1002.

(2) Hammond JS, Keeney RL, Raiffa H. Smart choices: a practical guide to making better decisions. Boston (MA): Harvard Business School Press; 1999.

(3) Raiffa H. Decision analysis: introductory lectures on choices under uncertainty. Reading (MA): Addison-Wesley; 1968.

(4) Merz JF, Fischhoff B. Informed consent does not mean rational consent. Cognitive limitations on decision-making. J Leg Med 1990:11:321–50.

(5) Fischhoff B. Eliciting knowledge for analytical representation. IEEE Trans Syst Man Cybern 1989;13:448–61.

(6) Morgan MG, Henrion M. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. New York (NY): Cambridge University Press; 1990.

(7) Clemen RT. Making hard decisions: an introduction to decision analysis. Boston (MA): PWS-Kent; 1991.

(8) Howard RA. Knowledge maps. Manage Sci 1989;35:903–22.

(9) Bostrom A, Fischhoff B, Morgan MG. Characterizing mental models of hazardous processes: a methodology and an application to radon. J Soc Issues 1992;48:85–100.

(10) Brokensha DW, Warren DM, Werner O. Indigenous knowledge: systems and development. Washington (DC): University Press of America; 1980.

(11) Budescu DF, Wallsten TS. Processing linguistic probabilities: general principles and empirical evidence. In: Busemeyer JR, Hastie R, Medin DL, editors. Decision making from a cognitive perspective. New York (NY): Academic Press; 1995. p. 275–318.

(12) Merz JF, Druzdzel MJ, Mazur DJ. Verbal expressions of probability in informed consent litigation. Med Decis Making 1991;11:273–81.

(13) Erev I, Cohen BL. Verbal versus numerical probabilities: efficiency, biases and the preference paradox. Organ Behav Hum Decis Processes 1990;45: 1–18.

(14) Fischhoff B, Bruine de Bruin W. Fifty/fifty = 50? J Behav Dec Making 1999;12:149–63.

(15) Poulton EC. Bias in quantifying judgment. Hillsdale (NJ): Lawrence Erlbaum; 1989.

(16) Fischhoff B. The real world: what good is it? Organ Behav Hum Decis Processes 1996;65:232–48.

(17) Quadrel MJ. Elicitation of adolescents' risk perceptions: qualitative and quantitative dimensions [dissertation]. Pittsburgh (PA): Carnegie Mellon University; 1990.

(18) McIntyre S, West P. What does the phrase "safer sex" mean to you? Understanding among Glaswegian 18 year olds in 1990. AIDS 1992;7: 121–6.
(19) Turner CF, Lessler JT, Gfroerer JC, editors. Survey measurement of drug use: methodological studies. Rockville (MD): National Institute on Drug Abuse; 1992.
(20) Leigh BC. In search of the Seven Dwarves: issues of measurement and meaning in alcohol expectancy research. Psychol Bull 1989;105: 361–73.
(21) Morwitz V. Behavioral intentions. In: Armstrong JS, editor. Handbook of forecasting principles. Norwell (MA): Kluwer; 1999.
(22) Fischhoff B, Slovic P, Lichtenstein S, Read S, Combs B. How safe is safe enough? A psychometric study of attitudes toward technological risks and benefits. Policy Sci 1978;8:127–52.
(23) National Research Council. Understanding risk: informing decisions in a democratic society. Washington (DC): National Academy Press; 1996.
(24) Slovic P. Perceptions of risk. Science 1987;236:280–5.
(25) Breyer SG. Breaking the vicious circle: toward effective risk regulation. Cambridge (MA): Harvard University Press; 1993.
(26) Wildavsky A. Searching for safety. London (U.K.): Transaction Books; 1988.
(27) Murphy AH, Lichtenstein S, Fischhoff B, Winkler RL. Misinterpretations of precipitation probability forecasts. Bull Am Meteorolog Soc 1980;61: 695–701.
(28) Kahneman D, Slovic P, Tversky A, editors. Judgments under uncertainty: heuristics and biases. New York (NY): Cambridge University Press; 1982.
(29) Plous S. The psychology of judgment and decision making. New York (NY): McGraw-Hill; 1993.
(30) von Winterfeldt D, Edwards W. Decision analysis and behavioral research. New York (NY): Cambridge University Press; 1986.
(31) Department of Health and Human Services, Food and Drug Administration. Over-the-counter human drugs; proposed labeling requirements. Federal Register 1997 Feb 27; 9024–62.
(32) National Research Council. Shopping for safety: providing consumer automotive safety information. Washington (DC): National Academy Press; 1996.
(33) Dawes R. Rational choice in an uncertain world. San Diego (CA): Harcourt Brace Jovanovich; 1988.
(34) Kahneman D, Tversky A. Choices, values, and frames. Am Psychol 1984; 39:341–50.
(35) Stone ER, Yates JF, Parker AM. Effects of numerical and graphical displays on professed risk-taking behavior. J Exp Psychol Applied 1997; 3:243–56.
(36) Linville PW, Fischer GW, Fischhoff B. AIDS risk perceptions and decision biases. In: Pryor JB, Reeder GD, editors. The social psychology of HIV infection. Hillsdale (NJ): Lawrence Erlbaum; 1993. p. 5–38.
(37) Slovic P, Fischhoff B, Lichtenstein S. Accident probabilities and seat-belt usage: a psychological perspective. Accident Analysis Prev 1978;10: 281–5.
(38) Schwartz LM, Woloshin S, Byram SJ, Sox HC, Fischhoff B. Which age is right? Women's reactions to the mammography screening debate [abstract]. J Gen Intern Med 1998;13(suppl):107.
(39) Morgan MG. Fields from electric power. Pittsburgh (PA): Carnegie Mellon University, Department of Engineering and Public Policy; 1995.
(40) Maharik M, Fischhoff B. Public views of using nuclear energy sources in space missions. Space Policy 1993;9:99–108.
(41) Maharik M, Fischhoff B. Contrasting perceptions of risks of using nuclear energy sources in space. J Environ Psychol 1993;13:243–50.
(42) Fischhoff B, Riley D, Kovacs D, Small M. What information belongs in a warning? A mental models approach. Psychol Marketing 1998;15: 663–86.
(43) Riley DM. Human factors in exposure analysis for consumer paint-stripper use [dissertation]. Pittsburgh (PA): Carnegie Mellon University; 1998.
(44) Fischhoff B, Downs J, Bruine de Bruin W. Adolescent vulnerability: a framework for behavioral interventions. Appl Prevent Psychol 1998;7: 77–94.
(45) Byram S. Breast cancer and mammogram screening: mental models and quantitative assessments of beliefs [dissertation]. Pittsburgh (PA): Carnegie Mellon University; 1998.
(46) Woloshin S, Schwartz LM, Byram S, Fischhoff B, Welch HG. Scales for assessing perceptions of event probability: a validation study. Med Dec Making 1998;14:490–503.

Improving Cancer Risk Communication: a Discussion of Fischhoff

Edward Maibach

The implications of Fischhoff's observations *(1)* about why cancer risk communication can be hard, and the steps risk communicators can take to be more effective, are rich at both the operational level as well as the conceptual level.

Operational Implications

Fischhoff's recommendations are clear and extremely promising. Following his recommendations will not be easy, however, as they represent a considerable challenge to the field of risk communication. Fischhoff recommends a formal set of procedures for capturing and comparing expert conceptions (an "expert model") and target audience conceptions (a "mental model") of the risk at issue. After these conceptions have been formalized, an iterative research-based process is used to determine which elements of the expert model most effectively improve target audience understanding of the risk. Whereas the skills to conduct these procedures may be stock in trade for risk and decision scientists, they are unknown to the majority of health and risk communication planners.

Accelerating the adoption of Fischhoff's formalized approach will require a considerable investment in research and development (R&D) and training, although if made, the dividends are likely to be considerable. R&D is needed to create simple yet robust means for developing expert and mental models, as well as empirical procedures by which to determine which information from the expert model is most useful. Health and risk communication professionals, students, and the people who teach and advise them will require training, at a minimum, to understand the rationale underlying this formalized approach. Until such time that risk communication professionals have a full understanding of the process, risk or decision scientists or both should be integral members of the communication planning team.

There are budgetary and timeline implications at the operational level as well. When properly conducted, communication planning requires an investment of time and money in audience or formative research. Fischhoff's formalized approach to risk communication planning, however, may pose even greater time and financial demands, at least initially. These demands may be lessened in the long run as expert models and audience mental models become available in the published literature and elsewhere.

Conceptual Implications

There are two distinct objectives for communicating cancer risk information: enabling audience members to make informed decisions (e.g., cancer patients who have a choice among treatment options) and persuading audience members to adopt a specific belief or behavior [e.g., increasing daily consumption of fruit and vegetables *(2)*]. Each of these objectives has a different underlying assumption. In the case of enabling people to make informed decisions, the communicator's assumption is that no one decision or course of action is right for all people, but rather each individual's unique values and situation must be weighed against the options. In the case of using risk information to persuade audience members to embrace a given option, the communicator's assumption is that there is an acceptably high degree of consensus among experts to warrant the assertion that the recommended option is the "right" option. Without a high degree of consensus that the recommended behavior is, in fact, in the best interest of the individual (or possibly, in certain instances, in the best interest of society), the ethics of providing risk information for purposes of persuasion are questionable *(3,4)*.

The criteria for judging effectiveness must, therefore, vary according to the objective of the risk communication program. Risk communication that seeks to enable people to make informed or "better" decisions should be judged on its ability to

- Impart an accurate (albeit perhaps limited) understanding of the risk, particularly (as Fischhoff points out) information that is "worth knowing";
- Facilitate a good decision-making process *(5)*;[1] and
- Impart confidence that a good decision has been reached. Without such confidence, decision-makers are likely to be plagued by unproductive anxiety *(6)*.

Conversely, risk communication that seeks to persuade should be judged by its ability to

- Impart an accurate understanding of the risk, particularly information that is motivational;
- Modify or reinforce cognitive antecedents to the recommended option; and
- Encourage positive self-evaluations on the basis of enactment of the recommended option *(7)*.

Fischhoff's formalized approach to assessment and communication planning have the potential to improve the practice and evaluation of both forms of risk communication. Risk communication planners should, however, be clear about which of the two objectives they are pursuing to apply the formalisms appropriately.

References

(1) Fischhoff B. Why (cancer) risk communication can be hard. J Natl Cancer Inst Monogr 1999;25:7–13.

(2) Maibach E, Holtgrave DR. Advances in public health communication. Ann Rev Public Health 1995;16:219–38.

Correspondence to: Edward Maibach, Ph.D., M.P.H., Porter Novelli, 1120 Connecticut Ave., N.W., Suite 1100, Washington, DC 20036 (e-mail: emaibach@porternovelli.com).

See "Note" following "References."

© Oxford University Press

(3) Rothschild ML. Carrots, sticks & promises: a conceptual framework for the management of public health and social issue behaviors. J Marketing. In press 1999.
(4) Smith C, Dunfee T, Ross W. Social contracts and marketing ethics. J Marketing. In press 1999.
(5) Mowen JC. Judgment calls. New York (NY): Simon & Schuster; 1993. p 161–85.
(6) Bandura A. Self-efficacy: the exercise of control. New York (NY): WH Freeman; 1997.
(7) Bandura A. Social foundations of thought and action: a social cognitive theory. Englewood Cliffs (NJ): Prentice-Hall; 1986.

Note

[1]A good "decision-making process" is distinct from a good "decision." A good decision-making process interprets risk information accurately, in the context of personal values and the situational considerations, to select the most highly valued option. The quality of a decision is more difficult to assess.

What Does It Mean to Understand a Risk? Evaluating Risk Comprehension

Neil D. Weinstein

Risk communications are frequently intended to help people understand hazards they face, with the hope that this understanding will help them make better decisions about the need for action or help them choose among alternative actions. To evaluate the success of such communications, a definition of "understanding" is needed. This paper suggests that decisions about personal risks require, at a minimum, information about the nature and likelihood of potential ill effects, information about the risk factors that modify one's susceptibility, and information about the ease or difficulty of avoiding harm. Even if these attributes are accepted as essential criteria for understanding, research on risk perceptions suggests that assessing what people know or believe is sometimes quite difficult. The focus of the paper is on the several dimensions of risk comprehension. Examples of how each can be assessed are drawn from research on public perceptions of the risks from smoking. These examples demonstrate that the public has only a limited understanding of smoking risks. [Monogr Natl Cancer Inst 1999;25:15–20]

Goals for Risk Communication

Many reasons exist why risk information may be communicated to the public or to individual patients: build trust, influence policy, fulfill legal obligations, deny responsibility, justify past actions, and more. Among the most important is the goal of helping people understand the risks they face. The expectation is that this understanding will help people better decide which hazards they should accept and which they should avoid or reduce.

But what does it mean to "understand a risk"? To determine how much the public understands about a hazard or to evaluate how effective a risk communication has been, an appropriate measure of understanding is needed. Hazards are complex and multidimensional, so a useful and defensible definition is far from obvious. Attempts have been made to establish criteria for assessing the success of a single-risk message *(1)*, to specify what policymakers need to know about a risk *(2)*, and to provide a comprehensive examination of the entire risk communication endeavor *(2,3)*, but no one has tried to specify the minimum information needed by an individual to understand his or her own risk. This paper will draw on the decision-making and health behavior literature to identify a small set of risk attributes that are essential knowledge for informed personal decision making, and it will draw on the risk perception literature to offer suggestions about how lay knowledge of these attributes might be assessed.

Despite this paper's focus on understanding, decisions and behavior are not determined by knowledge alone. Many other factors are involved and are sometimes much more powerful. These factors include emotions, personal values, social pressures, environmental barriers, and economic constraints. Educated individuals do not always make wise decisions. Still, without an understanding of the risk, appropriate decisions about personal action can come only from luck or from following someone else's advice or example, neither of which is a dependable guide.

Dimensions of Risk Comprehension

A comprehensive understanding of the risks from smoking would include such diverse topics as the identity of the chemicals in cigarette smoke, the transport and metabolism of these chemicals in the body, the economic costs of smoking, and the types of therapies used for smoking-related illnesses. If all of these topics are required for "understanding," few if any lay people can ever be expected to understand the risks of smoking. However, such complete knowledge is not needed to decide whether one should smoke. To make decisions about the importance of avoiding a particular hazard, such as smoking, people need, at a minimum, information about the nature and likelihood of potential ill effects, about the risk factors that modify their susceptibility, and about the ease or difficulty of avoiding harm.

A focus on these particular risk dimensions is consistent with a large body of theory and research on health behaviors *(4)* and decision making under uncertainty *(5–7)*. All of these theories explicitly consider the severity of the potential harm. That is, the theories assert that people are influenced by how serious they think it would be if the hazardous outcome occurred. A prereq-

Correspondence to: Neil D. Weinstein Ph.D., Department of Human Ecology, Rutgers University, 55 Dudley Rd., New Brunswick, NJ 08901-8520, (e-mail: neilw@aesop.rutgers.edu).

See "Notes" following "References."

© Oxford University Press

uisite for any accurate judgment of severity is obviously an understanding of the identity of the harm that might occur.

A second variable in relevant theories of decision making and behavior is the probability of the harm. In other words, what are the chances that it will occur if no preventive steps are taken? The probability of harm is rarely the same for everyone, so understanding one's personal vulnerability requires knowledge of the main factors—such as personal behaviors, family history, and environmental exposures—that determine that probability. "Controllability" is another concept found in theories of health behavior *(8)*. It denotes the reduction in the likelihood of harm that can be achieved by preventive action, so it can be viewed as one facet of understanding the probability of harm.

A final crucial topic is the ease or difficulty of carrying out actions to reduce the risk. Experience shows that people frequently underestimate the difficulty of carrying out preventive activities, so this topic should also be considered essential to informed decision making.

Many other risk dimensions might influence decisions and therefore might be considered necessary to understanding: slowly appearing versus rapidly appearing, familiar versus unfamiliar, vivid versus pallid, affecting many people simultaneously versus affecting individuals one at a time, natural versus man-made, and so on *(9)*. Because there is neither agreement nor much data about how influential these dimensions are, they do not seem to be essential to a core definition of risk understanding, although some might be very important in particular situations.

Clearly, no single question can ascertain whether people understand a risk. Only multiple queries will allow us to reach valid conclusions about the extent to which people understand a risk and the types of information they are lacking.

The remainder of this paper will elaborate on the risk dimensions just mentioned, including ways in which they might be assessed. Examples will be drawn from research on the public's views of smoking, including a recent review of the accuracy of smokers' risk perceptions *(10)*. Although the examples from the smoking literature provided here do not form a complete review of the smoking data, they nevertheless indicate that public knowledge—when compared to a comprehensive definition of understanding—is inadequate. Most people still do not understand the risks of smoking, and what they do know does not form an adequate basis for decision making.

Essential Criteria for Judging Comprehension

The Nature of the Potential Harm

Judging the severity of a hazard requires more than a vague understanding that an activity is "bad for you" or that it "causes cancer." Unless a person has a reasonably complete knowledge of the undesirable consequences of that activity—both what outcomes can occur and how serious these outcomes are—he or she is not in a position to decide how negative the consequences might be.

The Identity of the Health Outcomes That Can Occur

When surveys or interviews focus on a particular health problem, such as breast cancer, stroke, or acquired immunodeficiency syndrome (AIDS), the identity of the health threat is self-evident. However, research on risky behaviors and environmental hazards, such as smoking, alcohol abuse, or lead exposure, should not assume that people know what illnesses or disabilities may result.

Example: What are the potential health consequences of smoking? Can people generate a reasonably complete list of the illnesses that are caused by smoking? Do they acknowledge that smoking can cause these illnesses?

A major problem with most attempts to learn whether people know what illnesses are caused by smoking is that the correct answers are implicit within the questions. Respondents are asked whether smoking can cause heart disease, lung cancer, and so forth. The questions thus remind people of the health effects that are of concern and suggest what answer is correct. Such prompted recall is a much more lenient measure of knowledge than unprompted recall. Without being prompted in this way, it appears that many people cannot identify even the best-known, most serious illnesses caused by smoking. In two older studies cited in the 1989 Surgeon General's Report *(11)*, for example, only 32% of blacks surveyed and only 54% of a primarily white sample mentioned smoking when asked about the most important causes of cardiovascular disease.

It appears that for at least several decades, no survey in the United States has asked people, in an open-ended format, what illnesses can be caused by smoking. A 1990 Canadian survey *(12)* inquired, "To the best of your knowledge, what, if any, are the health hazards related to smoking?" The results were that only 44% of respondents cited lung cancer, only 20% cited heart disease, only 20% cited emphysema, and only one in 20, or less, cited other smoking-related illnesses such as high blood pressure, oral cancer, respiratory problems, bronchitis, or stroke. A weakness of this study is that prompts (e.g., "Can you think of anything else"?) were not used to encourage people to try to add to the list of illnesses that they had already mentioned.

If individuals cannot identify even the best known, most severe health effects of smoking without prompting, they are certainly unable to apply that information in deciding whether to smoke.

When questions ask about specific effects, lung cancer, heart disease, and emphysema are the diseases most commonly associated with smoking by the lay person. A large majority of survey respondents agree that cigarettes can cause those illnesses *(13)*. However, significant minorities of smokers do not agree. For example, according to the most current data reported in the 1989 Surgeon General's Report *(11)*, 29% of the smokers did not agree that smoking causes heart disease; 17%–25% did not agree that cigarette smoking causes lung cancer; and 21%–25% did not believe that smoking causes emphysema or chronic bronchitis. Similarly, in Australia, Chapman et al. *(14)* found that only about 50% of the smokers agreed that smoking can cause heart disease, poor circulation, bronchitis, and stroke, and only 72% agreed that smoking can cause lung cancer.

The Seriousness of These Outcomes

Perhaps because most health behavior research focuses on illnesses or risky behaviors that are serious and relatively familiar, it is assumed that lay people are well informed about the nature of these problems. Efforts are seldom made to determine what people know about the symptoms or duration of illness, the possible aversiveness of treatment, the success of therapies, and similar issues. Neither do researchers know which of these characteristics or others are the most important influences on health behaviors.

Example: How serious are the consequences of smoking? Are people aware of the difficulty or impossibility of curing these illnesses, of the severe ways in which smoking-induced chronic illnesses alter people's lives, and of the number of these illnesses that are potentially fatal?

Basic to understanding the severity of the consequences of smoking is appreciating what it is like to have lung cancer, heart disease, or emphysema and the chances of mortality from these diseases. Can people describe the debilitating symptoms of emphysema? Do people understand how deadly lung cancer is? Seldom are attempts made to answer such questions. The Canadian study cited earlier *(12)* found that people substantially underestimated the mortality rate from lung cancer, estimating on average that only 64.7% of cases of lung cancer result in death.

Some evidence shows that adolescents especially underestimate the severity of the harm from smoking. For example, adolescents gave lower ratings for the amount of harm caused by occasional, experimental, and regular smoking than did their parents *(15)*.

The studies just cited may convey a false impression that an individual's appreciation of the severity of an illness can be determined adequately from numeric ratings. Abundant evidence *(16)* shows that people are affected more by concrete examples and vivid images than by pallid statistics. Thus, knowledge of illness statistics—survival rates, illness durations, and so forth—surely fails to capture important aspects of illness awareness. Researchers might find ways to compare the perceptions of respondents with those of people who have actually experienced a disease, using the latter perceptions as a standard of accuracy.

The Probability of Harm

Assessing lay perceptions of probability is extremely challenging because members of the public do not appear to think in terms of rates, percentages, or odds, the units used by scientists to express event likelihood.

The Probabilities as Expressed in Absolute, Numeric Terms

A natural inclination is to ask people for numeric judgments of likelihood and to compare their responses with epidemiologic data. This approach conveys an impression of rigor and scientific objectivity, but conclusions about understanding reached in this way are highly questionable. Viscusi *(17,18)*, for example, implies that, when people are asked to give a numeric estimate of risk (e.g., the number of smokers out of 100 who will develop lung cancer or the likelihood of living to a certain age), the answers they give are meaningful to them and show whether people understand the risks of smoking. Such interpretations are simply not borne out by the data.

Lay people often have great difficulty understanding and using numeric estimates of risk. This fact is well known and well documented in the literature. For example, Lipkus et al. (personal communication, 1998) asked people, about 80% of whom had at least some college education, extremely simple questions regarding probabilities and percentages, with stunning results. Approximately 20% of the respondents (and about 40% of the smokers) were unable to correctly answer the following question: "If the chance of getting a disease is 10%, how many people on average would be expected to get the disease out of 100?" Yamagishi *(19)* found that people rated a cancer as riskier when it was described as killing "1286 out of 10 000 people" than when described as killing "24.14 out of 100 people."

Similarly, Black et al. *(20)* found that 38% of their generally well-educated sample (62% college graduates) made fundamental, logical errors in making numeric risk estimates. For example, some respondents gave a higher estimate for the likelihood of contracting breast cancer sometime in the next 10 years than for the likelihood of sometime in the next 20 years, whereas others gave a higher estimate for the likelihood of getting and dying of breast cancer than for the likelihood of getting breast cancer in the first place.

Data also indicate that people's actions sometimes bear little relation to statistics denoting absolute risk. A study by Shiloh and Saxe *(21)*, for example, found no correlation between the probabilities provided to couples during genetic counseling and the decisions they then made about childbearing. Windschitl and Wells *(22)* reported that judgments of likelihood using verbal categories predicted subsequent actions better than did numeric likelihood estimates. Consistent with that result, Diefenbach and Weinstein *(23)* found that college students said that scales of risk with verbal labels (e.g., "very unlikely") were easier to use and did a better job of representing their true feelings than did numeric scales using either odds or percentages.

All together, the available evidence shows that a large proportion of the public has enormous difficulty with odds and percentages. The ability to cite a statistic or make a numeric estimate does not mean that people understand what the number really means, that they use it in making decisions, or that they think it applies to them. Asking survey respondents to place a numeric probability on the occurrence of a health outcome and then comparing their answers with objective data is one of the least meaningful and least reliable measures of risk understanding.

Example: Can people correctly state the number of smokers out of 100 who will contract heart disease, lung cancer, or other smoking-related illnesses by a certain age? Can people correctly state the fraction of smokers who will prematurely die because of smoking? Can people correctly state the proportion of a specific illness, such as lung cancer, that is caused by smoking?

Borland *(24)* conducted several studies that used numeric risk assessments. In one, respondents were asked about the numeric risk of dying from three smoking-related diseases. When the risks of just those three diseases were added up, the total chance of dying was already more than a 100% (approximately 120%). Borland concluded that people do not respond to such questions with answers that obey the laws of probability.

Several studies that collected numeric probability estimates found a surprising number of people who estimated the risk of lung cancer from smoking as "50%" or as "50 out of 100"—as many as 37% of the respondents (Sutton SR: personal communication, 1995). Such a spike in the distribution of responses strongly suggests that people were using 50% as a sort of default answer, because they did not have any better notion of how to answer the question. It is likely that many of these respondents did not mean that the risk was actually one in two, but rather that the outcome might or might not happen.

The impossibility of using numeric risk estimates to decide whether people understand the risks of smoking is further demonstrated by the fact that studies forcing survey respondents to place numeric values on smoking outcomes yield inconsistent

answers about the accuracy of the public's perceptions *[see Table 1 in (10)]*. In some studies *(17,18)*, people appear to exaggerate the risks, estimating, for example, that about 40 smokers out of 100 eventually develop lung cancer, whereas a more accurate figure would be closer to 5% or 10%. Other studies suggest that people underestimate the risks. In a representative national sample in England *(25)*, the median response was that 100 out of 1000 smokers would die because of smoking before age 70 years. According to Sutton *(25)*, the epidemiologic data indicate that the actual number of deaths as a result of all smoking-induced illnesses is 250.

Even though lay answers to such questions are problematic, it might appear that the questions, at least, are clear and objective. This clarity, however, is illusory. The questions do not define "smoker" and leave ambiguous the number of cigarettes a day that are smoked or the number of years of smoking. When asked about a "smoker," most respondents will bring to mind a person who smokes rather heavily. The risks they estimate for an undefined smoker are probably higher than if they had been asked to consider the real world population of all smokers, which includes light smokers. Another flaw with such questions is the failure to ask people about their own risk, because we know that people seldom believe that the risks faced by others apply to themselves (see below).

The Relative Probability of the Hazard

If numeric measures of probability (e.g., the probability that a man 50 years of age will eventually develop prostate cancer) have little meaning for most people, how *do* they think about the likelihood of harm and how can we assess the accuracy of their thoughts? Do people use some set of verbal categories to think about likelihood? Do they think in terms of the absolute magnitude of individual risks, or do they relate more naturally to relative risks (i.e., which risks are bigger or smaller than others)? Do people even pay attention to absolute risks, or do they care more whether their own risks are higher or lower than the risks of their peers? As these questions suggest, there are many possibilities and even more assessment strategies. At present, there is no agreement concerning the answers to these questions.

An attractive alternative to judgments of absolute risk is to determine whether people have an accurate sense of the relative risk. Recognizing the public's difficulty with numeric estimates, relative risk should be assessed with rankings of different hazards or with qualitative comparisons (e.g., "the two hazards are 'about the same,'" or "hazard 1 is 'much greater' than hazard 2"), rather than with numeric estimates (e.g., "hazard 1 is 80% greater than hazard 2"). The following kinds of comparisons seem familiar and natural.

How does the probability of harm from the hazard of concern compare to the probability of harm from other familiar hazards? People deal with risks all the time, although not necessarily consciously. When making decisions about a hazard, it is helpful to consider how it compares with others for which one has already made decisions. This argument suggests that an adequate understanding of a risk includes an ability to place the risk into a real-life context in which a variety of hazards compete for attention.

Example: In relative terms, do smokers understand how the harm from smoking-induced illness compares with the harm from other familiar hazards, such as automobile fatalities, murder, and AIDS?

The evidence indicates that people do not understand how smoking compares with other risks. In an American Cancer Society study *(26)*, 28% of American voters believed that auto accidents kill more people than any other hazard on a list they saw that included cigarette smoking, drug use, AIDS, alcohol abuse, and murder. Only 21% recognized that cigarette smoking was the biggest killer on the list, although more deaths are attributable to cigarette smoking than to all the other listed health problems combined. Similarly, the 1989 Surgeon General's Report *(11)* cites a survey in which teenagers were asked to pick out the hazard that caused the most fatalities from a list that included traffic accidents, drug overdoses, fires, and cigarette smoking. The largest group, 44%, identified traffic accidents as the greatest hazard, followed by drug overdoses. Only 19% chose cigarette smoking. Similar underestimations of the relative risk of smoking have been reported by Borland *(24)* and Eiser et al. *(27)*.

How much is the probability of harm for someone at risk elevated above that of someone who is not in the same risk category? When the hazard in question is a particular risk factor, such as unsafe sex, a high-fat diet, or asbestos exposure, an obvious issue is the extent to which these risk factors raise the likelihood of victimization above that of people without these risk factors.

Example: In relative terms, do people realize how much the risk of smoking-related illnesses is increased above the risk of nonsmokers? Do they realize that smokers are substantially more likely to die prematurely? How do smokers' perceptions of the risks of smoking compare with nonsmokers' perceptions?

All the available studies indicate that the great majority of lay people realize that smoking is harmful and believe that the risk of diseases like emphysema and lung cancer is higher for smokers than nonsmokers *[see Table 2 in (10)]*. It is also found consistently, however, that smokers believe these health effects to be smaller and less well established than do nonsmokers *[see Table 3 in (10)]*.

Quite a few studies have asked smokers how their risk of becoming ill from smoking compares with the risk of the average person (or the average person their age). In a representative national sample, Ayanian and Cleary *(28)* found that 71% of smokers believed their personal risk of heart attacks to be average or below average when compared with the risk of other people of their age and sex, and 60% believed that their personal risk of cancer was average or below. In other studies, the mean comparative risk judgments of smokers ranged from "average" (i.e., smokers claimed that their own risk was no different than that of the average person) to "a bit higher" (Sutton SR: personal communication, 1995). In no case did smokers acknowledge that their risk of lung cancer, heart disease, or emphysema was "moderately," "substantially," or "much" higher than that of the average person. Because the epidemiologic data show that smokers' risk of getting these diseases is greatly elevated relative to nonsmokers *(11)*, for smokers to believe that their risk is hardly different from average demonstrates a clear failure of understanding.

Finally, Chapman et al. *(14)* found that smokers maintain a constellation of comforting, risk-minimizing beliefs. About one third of smokers, for example, agreed that "Many people who smoke all their lives live to a ripe old age, so smoking is not all that bad for you." For each of the 14 different risk-minimizing beliefs examined, smokers agreed more than did exsmokers.

What is the size of one's own risk compared with others with similar risk factors? Research has shown that risk perception is not an unbiased appraisal of information, but rather an attempt to seek the most comforting view of one's personal vulnerability that fits within the bounds of the evidence. As a result, whatever people believe about the risks faced by others, they tend to believe that their own personal risks are less *(29,30)*[1]. This phenomenon is robust and appears in many different groups, with various methodologies, and with nearly all hazards. Thus, the result of asking people about the likelihood that they personally will have a problem can be quite different from the result of asking them about people in general or about people their age.

This optimistic bias appears to be equally descriptive of adults and adolescents *(10,31)*. The magnitude of unrealistic optimism (i.e., the difference between risk estimates for oneself and for others) varies from hazard to hazard, but it is particularly large for problems, like lung cancer, that are believed to be preventable by individual action.

Example: In relative terms, how do smokers' views of their own risk compare with their views of the risk to other smokers? Do smokers realize that the risks of smoking apply to them, or do they believe that they are for some reason less at risk than other smokers?

Numerous studies *[see Table 4 in (10)]* demonstrate that smokers believe their risks from smoking are significantly lower than the risks for other smokers. Additional studies reveal a variety of mechanisms or rationalizations used by smokers to minimize their risk. For example, smokers' estimates of the number of years of smoking needed to produce adverse health effects increase with the number of years they have been smoking *(32)*. Smokers also claim that their cigarettes' tar content is lower and that their cigarettes are less hazardous than the average brand *(11,33)*.

The Factors That Influence Individual Susceptibility

Seldom is everyone's risk the same. People should be able to recognize when their risk is substantially higher or lower than the norm. Such moderating variables might include their behaviors, physiology, genetic inheritance, or environment. Not only is it important to determine whether people recognize the factors that alter their risk, but it is also equally important to discover the myths that they hold about risk factors. People tend to believe myths that help them feel that they are not at risk.

Example: How does the risk vary with the duration and amount of smoking? Does having a lifestyle that is otherwise healthy counteract the effects of smoking? Is it okay to smoke for a few years? Are there safe cigarettes?

Several studies provide data on this topic. Hahn et al. *(32)* found that among long-term smokers, those who smoked more cigarettes rated themselves as being at higher risk of developing smoker's cough than light smokers. However, among short-term smokers, there was no relationship between the amount they smoked and their perceived risk of illness. This finding is consistent with research by Slovic *(34)* showing that adolescents think that they can smoke for a while without risk. For example, 32% agreed that "there is really no risk at all for the first few years."

LoConte *(35)* explored agreement with a variety of misleading but comforting beliefs about risk-moderating factors. These factors included the ability to reduce one's risk with special patterns of smoking (smoking outdoors, periodically abstaining, smoking only part of each cigarette) or to counteract the risk with health-promoting behaviors (exercise, relaxation, diet). Consistently, smokers agreed more strongly with these thoughts than did nonsmokers.

The Difficulty of Avoiding the Harmful Consequences of the Hazard

Some precautionary behaviors—obtaining a vaccination, using seat belts, taking vitamins—are relatively easy to perform. Many others—losing weight, practicing safer sex, stopping smoking—are notoriously difficult. Knowledge about barriers to change can influence the willingness to be exposed to hazards for which arduous remedial measures might later be needed.

Example: Do people realize the addictive power of cigarettes? Especially at the time when they initially begin to smoke, do they think that they will be able to quit if they so choose?

For smoking, a key aspect of understanding is a recognition of the likelihood of becoming addicted. Studies demonstrate that adolescents underestimate the likelihood that they will get addicted to cigarettes and overestimate the likelihood that they will stop smoking. For example, the 1989 Surgeon General's Report *(11)* described two surveys which found that about one quarter of teenagers agreed with the statement, "Teenagers who smoke regularly can quit for good any time they like."

Cohn et al. *(15)* asked teenagers about their chances of encountering 19 problems—one of which was to "get hooked on cigarettes." The illusory belief that they were less likely to become addicted than their peers was greater than the optimism shown for any other problem, with the single exception of "getting hooked on drugs like marijuana." Teens were also more confident than their parents that they could escape addiction to cigarettes.

In a University of Michigan survey, "Monitoring the Future," high school seniors were asked, "Do you think you will be smoking cigarettes 5 years from now?" *[cited in (36)]*. Of seniors who smoked less than a cigarette per day, approximately 85% stated that they probably or definitely would not still be smoking after 5 years. When the same group was polled 5 years later, 58% had not quit smoking. Almost one third of seniors who smoked a pack a day thought that they, too, would quit within 5 years. After 5 years, only 13% had actually quit.

Conclusion

To assess someone's understanding of a health hazard is to study the constellation of beliefs that are relevant to decisions and behaviors concerning that hazard. Just as no single question can ascertain a person's understanding of the judicial system or a person's understanding of the human body, scientists must use a variety of questions and approaches to learn what people understand about a risk. A basic set of essential points would include the identity and severity of the potential harm, the likelihood of harm under various circumstances, and the possibility and difficulty of reducing that harm. To make a decision, the same types of information would be needed about any other alternatives under consideration. Most current studies of hazard perception or knowledge fail to examine even this limited range of risk dimensions and use methodologies that are questionable. By identifying a small number of core issues and by pointing out some problems in their assessment, it is hoped that this paper will encourage researchers and communicators to pay close at-

tention to the kinds of information people need to make appropriate risk decisions.

References

(1) Sandman PM, Weinstein ND. Some criteria for evaluating risk messages. Risk Anal 1993;13:103–14.

(2) National Research Council. Improving risk communication. Washington (DC): National Academy Press; 1989.

(3) Stern PC, Fineberg HV, editors. Understanding risk: informing decisions in a democratic society. Washington (DC): National Academy Press; 1996.

(4) Conner M, Norman P, editors. Predicting health behaviour: research and practice with social cognition models. Buckingham (UK): Open University Press; 1996.

(5) Camerer C. Recent tests of generalizations of expected utility theory. In: Edwards W, editor. Utility theories: measurements and applications. Boston (MA): Kluwer Academic Publishers; 1992. p. 207–51.

(6) Edwards W. The theory of decision making. Psychol Bull 1954;51: 380–417.

(7) Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979;47:263–91.

(8) Ajzen I, Madden TJ. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. J Exp Soc Psychol 1986;22: 453–74.

(9) Slovic P. Perception of risk. Science 1987;236:280–5.

(10) Weinstein ND. Accuracy of smokers' risk perceptions. Ann Behav Med 1998;20:135–40.

(11) U.S. Department of Health and Human Services (USDHHS), Public Health Service. Reducing the health consequences of smoking—25 years of progress. A report of the Surgeon General. DHHS Publ No. (CDC) 89-8411. Washington (DC): U.S. Gov Print Off; 1989.

(12) Environics Research Group Ltd. Awareness of health hazards due to smoking. Toronto (ON): Environics Research Group Ltd; Dec 17, 1990.

(13) Gallup Organization. Gallup poll, July 6–8, 1990. Princeton (NJ): Gallup Organization; July 1990.

(14) Chapman S, Wong WL, Smith W. Self-exempting beliefs about smoking and health: differences between smokers and ex-smokers. Am J Pub Health 1993;83:215–9.

(15) Cohn LD, Macfarlane S, Yanez C, Imai WK. Risk-perception: differences between adolescents and adults. Health Psychol 1995;14:217–22.

(16) Nisbett R, Ross L. Human inference: strategies and shortcomings of social judgment. Englewood Cliffs (NJ): Prentice Hall; 1980.

(17) Viscusi VK. Do smokers underestimate risks? J Pol Economy 1990;98: 1253–69.

(18) Viscusi WK. Smoking: making the risky decision. New York (NY): Oxford University Press; 1992.

(19) Yamagishi K. When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. Appl Cogn Psychol 1997;11: 495–506.

(20) Black WC, Nease RF Jr, Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst 1995;87:720–31.

(21) Shiloh S, Saxe L. Perception of risk in genetic counseling. Psychol Health 1989;3:45–61.

(22) Windschitl PD, Wells GL. Measuring psychological uncertainty: verbal versus numeric methods. J Exp Psychol 1996;2:343–64.

(23) Diefenbach MA, Weinstein ND, O'Reilly J. Scales for assessing perceptions of health hazard susceptibility. Health Ed Res 1993;8:181–92.

(24) Borland R. What do people's estimates of smoking related risk mean. Psychol Health 1997;12:513–21.

(25) Sutton S. How ordinary people in Great Britain perceive the health risks of smoking. J Epidemiol Community Health 1998;52:338–9.

(26) American Cancer Society. Survey of U.S. voter attitudes toward cigarette smoking. Washington (DC): American Cancer Society; Sept 9, 1993.

(27) Eiser JR, Sutton SR, Wober M. Smoking, seat-belts, and beliefs about health. Addict Behav 1979;4:331–8.

(28) Ayanian JZ, Cleary PD. Perceived risks of heart disease and cancer among cigarette smokers. JAMA 1999;281:1019–21.

(29) Hoorens, V. Unrealistic optimism in health and safety risks. In: Rutter LD, Quine L, editors. The social psychology of health and safety: European perspectives. Aldershot (UK): Avebury; 1994. p. 153–74.

(30) Weinstein ND, Klein WM, editors. Unrealistic optimism. Special issue of J Soc Clin Psychol 1996;15.

(31) Quadrel MJ, Fischhoff B, Davis W. Adolescent (in)vulnerability. Am Psychol 1993;48:102–16.

(32) Hahn A, Renner B, Schwarzer R. Perception of health risk: how smoker status affects defensive optimism. Anxiety Stress Coping 1998;11: 93–112.

(33) Segerstrom SC, McCarthy WJ, Caskey NH. Optimistic bias among cigarette smokers. J Appl Soc Psychol 1993;23:1606–18.

(34) Slovic P. What does it mean to know a risk? Adolescents' perceptions of short-term and long-term consequences of smoking. J Behav Decis Making. In press 1999.

(35) LoConte JS. Optimistic risk perceptions and external risk attributions among smokers: identifying dimensions of risk-minimization and their correlates [dissertation]. New Brunswick (NJ): Rutgers University; 1995.

(36) Lynch BS, Bonnie RJ, editors. Growing up tobacco free: preventing nicotine addiction in children and youths. Washington (DC): Institute of Medicine; 1994.

Notes

[1]Unpublished bibliography on perceived invulnerability and optimistic biases about risk or future life events is available from the author.

I thank Mark Eichorn for his assistance during the preparation of this paper and Caron Chess for her helpful comments.

Introduction of Section: Overarching Considerations in Risk Communications: Romancing the Message

Joann E. Rodgers

If evaluation of factual information, by itself, informed the debate each of us has with ourselves about risky behaviors, this would have long ago been a nation full of nonsmoking, seat belt-wearing, lean, exercising, chemopreventing, sun-avoiding, health care-seeking men and women. That we are not such a nation also is underscored by the American experience with Alar, silicone breast implants, and screening mammography. They stand as powerful testimony not only to the influence and impact of the media on risk perception and behavior but also, most of all, perhaps, to the difficulty risk communicators face in crafting messages that individuals find meaningful *(1)*.

Put another way, individual perception and assessment of risk is fundamentally a subjective enterprise. Groups do not decide to quit smoking. Individuals do. And emotional, temperamental, and life-history factors—along with factual information—guide the decisions we make—or do not make—about the information we seek or get. One person's adventure is another's mortal threat. She calls racing cars great fun; he calls it a mobile death wish. He cares little for annual prostate-specific antigen tests; she demands screening mammography at age 35 years. Consequently, whereas it may be popular in scientific circles, whining about "irrational" and "unscientific" public assessments of risk not only gets us nowhere, but it also threatens to turn off interest among the very people we most want to better inform about risks.

Communications and behavioral scientists have long demonstrated that the most effective way to teach, persuade, or attract attention is with one-on-one face time. Doctors talking to patients patiently, helping each one understand the risks and benefits of this or that decision. Teachers mentoring students intensively. Parents explaining something multiple times when their children are ready to listen. And, in the best of all possible worlds, and with unlimited resources, that is what we know would work best for communicating risk.

Realistically, however, in the risk communications business, one-on-one time, because it is resource intensive, must be selectively used; most of the teaching, persuading, and attracting of attention must be left to collective forms of communications and mass media. The mass communications perspective is robust but presents challenges of its own.

The good news is that mass media, especially, can be powerful and cost-effective venues for health risk messages. According to a 1997 Kaiser Family Foundation *(2)*, three fourths of adult men and women consider magazines to be an "important" source of information on a variety of health topics, including breast cancer, pregnancy, human immunodeficiency virus or acquired immunodeficiency syndrome, and abortion. Results of a National Health Council (NHC) survey *(3)* reported earlier this year found that the media are the leading source for medical and health information among healthy 18- to 59-year-olds, although among those with chronic conditions, physicians are the primary source, followed by television news magazines and print news. Moreover, 53% of Americans, the NHC survey reported, say they have discussed a medical or health news story they saw or read with their doctor and that their doctors were mostly happy to do so. The survey found that 82% of the respondents believe health and medical news in mass media help them lead a healthy life and *58% report changing their behavior or having taken some type of action as a result of a medical news story* (italics mine). Seventy-six percent said they took the advice offered in the story, particularly if it had to do with avoiding a certain food, increasing exercise, talking to a doctor, reducing or stopping smoking, and dieting.

The bad news is that there appears to be a disconnect between the value scientists and doctors place on the information disseminated by mass media and the value placed on it by viewers and readers. Doctors routinely label a significant portion of health and medical coverage sensationalized, incomplete, inaccurate, unbalanced, or misleading, whereas the vast majority of Americans find it highly credible. In the NHC survey, for instance, 87% of Americans consider *60 Minutes* and *20/20* the most believable media sources, followed closely by television news in general (85%), daily newspapers (83%), and radio news (81%).

The worse news is that individuals come to mass media information and risk messages with an impoverished level of science literacy. A Dartmouth survey *(4)* of women 40–50 years old found they overestimated by 2000% their chances of dying of breast cancer within 10 years and exaggerated the relative reduction of risk by mammography screening by 600%. Also, they overestimated the absolute risk by 10 000%, according to Black et al. *(4)* And the still worse news: The women in the study had higher incomes and levels of formal education than the general population—91 of 145 had college or postgraduate degrees.

Other surveys have found that more than half of Americans incorrectly believe breast cancer is the leading cause of death for women, that four in 10 think spicy food causes and think Alar causes cancer, and that one third think eating sweets causes diabetes. In a landmark paper, Miller *(5)* reported that, in 1990, only 36% of American adults met even the most basic standards for scientific literacy as defined by a nine-item true or false quiz. On one question, for example, only 47% correctly disagreed that the earliest humans lived at the same time as dinosaurs. In a follow-up publication in 1998 in the National Science Board's *Science and Engineering Indicators,* Miller and Kimmel *(6)* reported that only 11% of Americans can define the word "molecule" and that only 20% understand what "DNA" means. Moreover, he wrote, only about one in three Americans understands the foundations of scientific inquiry—e.g., how theories are developed, how controlled experiments are designed, and how valid comparisons are made.

Additional complications and challenges emerge from both the routine and innovative activities of institutions and their

Correspondence to: Joann E. Rodgers, M.S., Media Relations, Office of Communications and Public Affairs, The Johns Hopkins Medical Institutions, 550 N. Broadway, Suite 1100, Baltimore, MD 21205-2011 (e-mail: jrodgers@jhmi.edu).

© Oxford University Press

public information and public relations staffs (Adams WC: personal communication). In their legitimate quest for attention, loyalty, funds, patients, and students, they do even less than the press to follow up or bring closure to previous breathless reports of "promising" advances or new discoveries that have an impact on risk perception. Does their work serve the cause of good risk communication? Or does it merely increase the noise-to-signal ratio by spewing out a constant stream of "promising solutions" to "medical problems" and risks in need of "cures" and "reductions" that involve tests, screening examinations, drugs, and expensive procedures. In press-release land, almost everything is a mortal threat.

How then to bridge this multiple credibility gap, lower the volume, clean out the confusion, and improve the process so that it works without resorting to propaganda or repeal of the First Amendment?

Sociologist Dorothy Nelkin has pointed out in numerous studies and books that the effect of messages delivered via mass media depends heavily on the social context in which the messages are received. In a 1996 article in the *Lancet (7),* she wrote that social context ". . . may include the readers' personal experience or prior knowledge, and the cumulative influence of previous media reports, popular representations and other sources of information . . . such as the television doctor stories." Seeking to trade on the impact of the latter, in fact, public affairs specialists at The Johns Hopkins Medical Institutions have persuaded the two top-rated "doctor shows" on television—*Chicago Hope* and *ER*—to tie in preproduced consumer health "programettes" designed to trade on the entertainment value of the shows to slip in presumably useful information. The concept behind "Living With Hope" and "Following ER" is enticing, and, in fact, it is successful, pulling in foundation grants and commercial sponsors. But whether or not this is a good use of resources, or even a credible form of public health education, is an unanswered question. Neither Johns Hopkins, nor the television producers, nor the funding agencies and sponsors have any idea why the spots might be effective, or whether anyone in the viewing audience finds the messages credible or merely another blob of "infotainment." The medium is powerful all right, but like a windmill's energy, sounds and images need a harness and a fixed channel to be useful. Some efforts are under way to evaluate the effect of "Following ER." The program's utility as a model depends heavily on sorting out what viewers are willing—or able—to bring to and take away from the experience.

Summarizing the overall focus of the panel devoted to overarching considerations in risk communications, it is helpful to mutate a phrase from recent presidential election politics: It's the context, stupid, along with the noise-to-signal ratio of risk messages. As with the economy, risk communication is a complex social science, not always amenable to exactitude and metrics. Thus, crafting and distributing messages that reach to the core of people's inner debate and grasp at behavior change remains an art as well as a science.

The experts gathered for this panel addressed some of these contextual considerations with novel insights into the process of communicating risks. Out of their remarks, it is hoped, will emerge some tips, based on their experience, to improve the process. In earlier conversations with them, all hummed a persistent leitmotif: the need to embrace the complexity of it all—to romance the message, if you will—if we wish to hit target audiences and make an impact.

In a world of limited resources, there are not enough means—even if we could figure out the ways—to reach every audience with every message. Priorities must be set, always a public policy challenge. Assuming key audiences are identified, audience attitudes, beliefs, behavior patterns, and values must be sorted out, along with what matters or does not matter to them as groups and as individuals. Cultural patterns, sex, age, class, and race must be figured into the sort, along with "influencers" that cut across all cultures, sexes, ages, classes, and races.

Fears must be not only correctly unveiled but also approached with caution. Women, for example, largely report breast cancer, or cancer in general, as their biggest health fear, although heart disease kills twice as many of them with symptoms and consequences every bit as crippling as cancer. But from a social and health policy standpoint, do risk communicators want to leave the two disease factions to duke it out with competing and conflicting risk messages or risk trivializing one risk or another?

Matthew Kreuter, an authority on health communication research at Saint Louis University, suggests that what little research there is in cancer risk communication has failed to account for how individuals with multiple and sometimes conflicting risks differ in their risk perceptions and behaviors, a failure with profound consequences for message crafters.

Elmer Huerta, a Peruvian-born physician with a long track record in using mass communication, especially radio, to inform the Hispanic community about health and health risks, and the founder of a nonprofit company that produces and distributes educational material for the Latino community in the United States, argues passionately for culture-based risk communications. In the Latino world, for example, failure to understand the particular bonds encompassed in the word "familismo" may well doom risk messages to the communications garbage heap.

Alfred Marcus, a behavioral scientist at the AMC Research Center in Denver, CO, combines an interest in communication research with practical experience in developing communication intervention messages for widespread use by managed care organizations and special populations. His remarks focus in part on identifying potential areas of research that will improve risk communications.

References

(1) Stocking H. How journalists deal with uncertainty in science. In: Friedman S, Dunwoody S, Rogers C, editors. Communicating uncertainty: media coverage of new and controversial science. Mahwah (NJ): Lawrence Erlbaum Associates; 1999.

(2) Treise D, Walsh-Childers K. 1997 Kaiser Family Foundation survey results. Presented at Conference for Magazine Editors and Writers sponsored by Columbia University Graduate School of Journalism and Kaiser Family Foundation; 1997 Apr 18.

(3) National Health Council. 21st century housecall: the link between medicine and the media [online] [cited 1998 Feb] Available from: http://www.healthanswers.com/Sourc.../21st_century/21st_housecall-2.htm

(4) Black W, Nease RF Jr, Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst 1995;87:720–31.

(5) Miller J. The public understanding of science in the United States, 1990. Presented at annual meeting of the American Association for the Advancement of Science; 1991 Feb 15; DeKalb (IL): Public Opinion Laboratory.

(6) Miller J, Kimmel L. Science and technology: public attitudes and public understanding. In: National Science Board. Science and engineering indicators, 1998. Chicago (IL): Chicago Academy of Sciences; 1998.

(7) Nelkin D. An uneasy relationship: the tensions between medicine and the media. Lancet 1996;347:1600–3.

Communicating Health Risk to Ethnic Groups: Reaching Hispanics as a Case Study

Elmer E. Huerta, Everly Macario

United States: a Multicultural Society

The United States is increasingly becoming a culturally heterogeneous society. In 1990, 75% of the people in the United States were white of European descent, 12.3% were African-American, 9.0% were Hispanic-American, 2.8% were Asian-American, and 0.7% were Native-American. It is projected that by the year 2021, these proportions will change: only 53% of the U.S. population will be white of European descent, the number of Asian-Americans and Hispanic-Americans will triple, and the number of African-Americans will double *(1)*.

Health Risk Messages: Not Effectively Reaching Ethnic Groups

Adequate understanding of health risks by the public can help to reduce the incidence and mortality of certain diseases, including cancer. The process of risk communication, however, has been studied infrequently with minority and lower-income communities, even though they may be at greatest risk *(2,3)*. Epidemiologic research has long shown that ethnic minorities suffer disproportionately from illness and death *(4)*. One reason for this gap may be that pertinent health messages have not reached ethnic groups effectively *(5)*. As many health professionals corroborate, communicating health messages in a culturally appropriate and sensitive manner (regardless of the specific health issue or disease in question) has become a challenge that can no longer be ignored. One way to begin addressing this issue is to move beyond the simple, direct translation of health messages already being used with the mainstream (white) population to messages that are culturally as well as linguistically appropriate *(6)*.

Defining Culture

Social scientists have been intrigued with the concept of culture and with the challenge of defining "culture" for many years *(7)*. "Culture" is a system of shared beliefs, customs, values, and behaviors that members of a society transmit from generation to generation and use to cope with one another and their world *(3)*.

To communicate health risk information that is culturally appropriate, it is critical to understand that people are the product of past experiences, cultural beliefs, and cultural norms and are thus culturally unique. It is also important to remember that there is as much diversity within cultural groups as there is across cultural groups *(8,9)*.

Defining Ethnicity

Ethnicity, conceptually, resembles the concept of culture in that it is collectively created over the course of generations out of the shared experience of a group of people who come to identify themselves as a group with unique history and origins *(7)*. Confusion of the term ethnicity with the constructs of culture stems from the fact that ethnicity is generally used to represent at least three distinct constructs: 1) "ethnicity" or the collective culture of a minority cultural group with some distinctive cultural characteristics within a larger society, 2) "ethnic origin" or a classification system on the basis of one's biologic ancestors, and 3) "ethnic identity" or the extent to which an individual chooses to incorporate a particular ethnic classification into his or her sense of self *(7)*.

The term ethnicity, therefore, is a multiplex concept with a variety of possible contributing attributes (e.g., intermarriage, self-identification, migration, level of acculturation, political and religious ties to the ethnic group, place of residence, socioeconomic status, and educational achievement) *(10)*. Which ethnicity becomes the dominant culture at home when an individual from one ethnic group marries an individual from another ethnic group? Does the male's or the female's ethnic background define the household? What about the children—how will their upbringing shape their views of health and disease? Try defining your own ethnicity—a formidable challenge you will more than likely discover. What traditional customs, passed through several generations, did you experience during your upbringing? Do you use home remedies that your ancestors used generations ago? What beliefs about health and cancer were translated to you from generations before?

Ethnicity remains a vital force in the United States. But, for many years, in striving to have a strong identity, the United States fostered the notion of the "melting pot," the idea that ethnic group distinctions are irrelevant and can be subsumed by a common American national identity. However, we have not "melted"—there is increasing evidence that ethnic values and identification are retained for many generations after immigration *(11)*. Rather than a "melting pot," the United States is more a mosaic or garden salad where each of our cultural differences retains its integrity while coexisting contiguously.

Distinct Ethnic Values or the Danger of Stereotyping

Although there are important sociodemographic differences among individuals in ethnic groups (according to factors, such as country of origin, educational attainment, and acculturation level), ethnic groups (e.g., Hispanics or Asians) seem to share core cultural values that differentiate them from other groups *(6)*. Whereas a danger of stereotyping exists, knowledge of an

Affiliations of authors: E. E. Huerta, Cancer Risk Assessment and Screening Center, Washington Hospital Center, Washington, DC; E. Macario, Porter Novelli, Washington, DC.

Correspondence to: Elmer E. Huerta, M.D., M.P.H., Cancer Risk Assessment and Screening Center, Washington Hospital Center, 110 Irving St., N.W., Rm. C1179, Washington, DC 20010.

© Oxford University Press

ethnic group's core values is an excellent starting point for successful health risk communication *(6,12)*. Despite the danger of stereotyping, it plays an important role in how we communicate—our communication depends in part on how we characterize the recipients of our messages *(12)*.

To be successful with Hispanics, for example, health communication messages need to consider the following core Hispanic values: "familialism" (the significance of the family to the individual), "collectivism" (the importance of friends and extended family in helping to solve problems), "simpatia" (the need for smooth interpersonal relationships in which criticism and confrontation are discouraged), "personalismo" (the preference for relationships with members of the in-group), "respeto" (the need to maintain one's personal integrity and allow for face-saving strategies), and "power distance" (certain persons, such as the powerful, the elderly, and the educated, should be treated with special deference) *(6)*. Keller and Stevens *(3)* have also discussed the importance of using critical family and kinship members to support and encourage health-promoting behavior among Hispanics. Johnson and Delgado *(8)* agree on the importance of family among Hispanics but recommend avoiding superficial references to family and, instead, encourage the identification of powerful images and messages that can activate positive behaviors directly linked to prevention goals.

"Collectivist" Versus "Individualist" Cultures

Cultures differ in the extent to which they pay attention to context and in the way communication takes place. Differences in social behavior can be found between "collectivist" and "individualistic" cultures *(13)*. When communicating health risk information, it is important to consider which of these two types of culture you are trying to reach.

People in collectivist cultures pay more attention to context when they communicate than do people from individualistic cultures *(13)*. Collectivists are not as explicit, direct, or clear as the individualists. By paying attention to the paralinguistics (emotional expression, gestures, touching, distance between bodies, body orientation, level of voice, and eye contact) but ignoring the verbal message, people from some cultures can easily misunderstand what a person from an individualistic culture is saying *(13)*. Time orientation, perceptions of personal space, exposure to private body parts, and discussion of intimate issues have unique variability across cultures *(3)*.

In individualistic (low-context) cultures, people distrust what is not said clearly; the best arguments are presented first to attract attention and generate a desire to hear the whole argument *(13)*. In contrast, the argument is presented climactically in collectivist (high-context) cultures—the messenger starts with peripheral arguments and gets to the punch line only after the other person has had time to react to and "digest" the information *(13)*. Differences between these two types of culture have implications for the impact of health communication messages.

Defining Acculturation

Recent immigrants are ones who have arrived in the United States during the past 10–15 years. Are they people who have been acculturated to the U.S. way of life? It depends. Acculturation is "the process by which the collective culture of a group is changed through contact with another culture" *(7)* or the degree to which members have immersed themselves in the culture of the country in which they live. Some second-, third-, or fourth-generation Americans differ from the dominant culture in values, lifestyle, and behavior and still consider themselves to be immigrants by the simple fact that they are "descendants of immigrants."

Whether or not a family lives in an ethnic neighborhood will influence the extent to which a family's cultural heritage will be preserved, such as the types of food consumed. This phenomenon can be observed by the "McDonaldization" process. In immigrants' home countries, grains, legumes, vegetables, and fruits may have been the staple diet; on arrival to the United States, these nutritious foods are often replaced with hamburgers, hot dogs, French fries, and soda. This transition occurs in part because this type of "fast food" is 1) the "American" way, 2) inexpensive, and 3) quick. Even though recently arrived immigrants who have higher socioeconomic and educational levels tend to behave more like the dominant culture, they do not, ironically, engage in healthier lifestyles necessarily.

Palinkas et al. *(14)* found that cessation and prevention programs that target Hispanic populations must be especially sensitive to sex-specific acculturation processes that create a discrepancy between the desire to adopt the lifestyles of the dominant culture and the recognition of the concomitant health-related risks. The results of the study by Palinkas et al. provided evidence for the argument that health-related behaviors and rates of disease begin to approximate those of the dominant society with increasing acculturation, despite their potential harmful consequences. Although the prevalence of smoking in the United States is apparently in decline, the majority of Hispanics (who possess relatively low socioeconomic status [SES] and work mainly in blue-collar occupations) adopt the smoking behaviors of non-Hispanic whites of similar SES, who have the highest rates of current cigarette smoking in the general population. The study *(14)* showed that Hispanic women are especially likely to adopt these smoking behaviors with increasing acculturation because cigarette smoking symbolizes the acquisition of greater status equality associated with changing sex roles and employment opportunities as well as the abandonment of Hispanic cultural sanctions and social pressures, which traditionally worked to prevent cigarette smoking in women but not in men.

Ethnicity and Perceptions of Health Risk

The level at which ethnic groups have severed ties with their country of origin plays a role in determining how they feel about life, death, and illness *(11)*. Studies *(11)* have shown that ethnic groups differ in 1) their reports of pain, 2) what they label as a symptom, 3) how they communicate about their pain or symptoms, 4) their beliefs about the cause of their illness, 5) their attitudes toward helpers (doctors and therapists), and 6) what treatment they desire or expect.

Hispanic women, for example, are more likely than non-Hispanic women to fear cancer and have fatalistic attitudes toward the disease (e.g., that illness is a matter of chance or fate and that fast recovery from illness is a matter of good luck) *(15)*. Roche et al. *(16)* found that cancer terms (e.g., "yearly," "mammogram," "diagnosed," "risk factors," "at risk," and "cancer") mean different things to English speakers (non-Hispanic women) and Spanish speakers (Hispanic women). For example, more than four times as many women interviewed in English as women interviewed in Spanish correctly defined "at risk." Incorrect responses commonly implied that you are at risk if you

do not have check ups. Thirty-two percent of the women interviewed in English in the study by Roche et al. *(16)* study understood "risk factors" as opposed to only 13% of the women interviewed in Spanish. Incorrect Spanish responses included defining "risk factors" as having cancer or as being ill and not taking care of yourself.

Place of residence may influence knowledge, attitudes, and behavior toward health and disease. Previous studies have shown that Hispanic women underuse cancer screening services. Morgan et al. *(17)*, however, found that urban Hispanic women engage in relatively high levels of cancer preventive behaviors, even though they tend to have a limited knowledge base of cancer coupled with misperceptions of cancer (e.g., beliefs that bumps and bruises cause cancer and that surgery causes cancer to spread).

"Risk Trade-off" Scale

Given that life events differ so widely among ethnic groups, how is the concept of health risk perceived? People have a "risk trade-off" scale: On one side of the scale is what to do; on the other side is what not to do. Both the health information one receives and one's cultural beliefs are factors that determine the weight of each side of the scale. The middle of the scale marks the behavior in which an individual engages, or how an individual negotiates the relative weight of each side of the scale. This trade-off, or balancing of the two sides of the scale, is essential to understanding and realizing that health decisions change depending on the cultural group to which an individual belongs and consequently on the unique distribution of weight on each side of the scale.

Importance of a Trusted Messenger

For a messenger to successfully communicate a health message to a receptor, the messenger must be familiar with the subjective culture of the people with whom he or she communicates. Accuracy implies that both the source and the recipient of a message assign the same meaning to it *(13)*. Most important, the messenger needs to be trusted by that community, and the cultural group should identify with him or her. A survey *(5)* sponsored by the Morehouse School of Medicine and the New America Wellness Group found that 61% of Hispanics, 28% of blacks, and 11% of whites said seeing a doctor of the same ethnicity was important to them.

Case Study: the "Cuidando su Salud" Model

Dr. Elmer Huerta's case study describing how mass media are used to reach Hispanics successfully will illustrate some of the aforementioned points. Kevin Kose, President, National Public Radio, once stated: "Radio, of all the media, is the most ephemeral and powerful force for transmitting ideas. It's a human voice from a remote location to another human being. And it's ubiquitous." Radios are especially integral to the Hispanic culture; they are kept in one's pocket and, therefore, are a personal companion. In Latin America, birthdays, deaths, and other family and community announcements or events are communicated via the radio. In 1994, 87.9% of Hispanics in the United States listened to Spanish-language radio on a daily basis. Hispanic adults 18 years or older listen to more than 2 hours of Spanish-language radio each day, on average *(18)*.

About 78% of Hispanic-Americans more than 5 years of age speak a language other than English at home. Nearly 49% of Hispanics in 1991 reported that they spoke English very well, whereas 14% said they spoke English well, and 37% said they spoke it poorly or not at all *(19)*. Because ethnic identity is inseparable from the language of the members of a group, health messages transmitted via the radio will be more successful if they are in the native language of the ethnic group *(3,20,21)*. Spanish-language radio should, therefore, be considered as a medium for reaching Hispanics with health-risk messages (especially for the more linguistically isolated Hispanics who may not be exposed to messages transmitted through other channels) *(6)*.

In large urban areas, Hispanics often depend on local Spanish-language broadcast and print media for news and entertainment [this is important because, among other reasons, urban Hispanic women tend to have inadequate levels of knowledge and misperceptions about cancer *(17)*]. Since 1989, Dr. Elmer Huerta's radio show, "Cuidando su Salud" ("Taking Care of Your Health"), has provided daily disease-prevention and health-promotion messages to Hispanics on a Washington, DC-area Spanish-language station. The show lasts 2 minutes, is aired three times a day during news segments, and has recently been expanded to reach more than 70 radio stations across the United States and Puerto Rico. By April 1999, 1847 different individual programs had been produced. Using Healthy People 2000 guidelines, these programs included 717 on health promotion (including programs on tobacco, nutrition, fitness, family planning, alcohol, violence, and mental health), 210 on health protection (including programs on unintentional injuries, food and drug safety, and environmental health), 875 on preventive services (including cancer, acquired immune deficiency syndrome, infectious diseases, maternal and infant health, and policy issues affecting Hispanics), and 45 on other health topics.

A synergy of health-message placement is created by expanding the channels to television, print, and the Internet. Not only are Dr. Huerta's Spanish-language radio vignettes aired daily, but health messages are also discussed during his daily radio talk shows and television show on Saturdays. A quarterly newsletter and information on the Internet further extend the reach to the community.

Dr. Huerta's radio show has been used successfully to recruit people for research projects *(22)* and cancer clinical trials *(23,24)*. The radio component of Maryland's Centers for Disease Control and Prevention-funded Breast and Cervical Cancer Campaign involved a single program that was aired three times 1 week apart in 1993. By June 1994, 655 women had been screened, 44% of whom were Hispanic. This ethnic breakdown of women screened contrasts sharply with the proportions of Hispanics in the two participating Maryland counties: 7% in Montgomery County and 4% in Prince George's County. At least one third of the Hispanic women attending the screening program had learned about it from "Cuidando su Salud" *(23)*. More than 300 Hispanic women were recruited for the National Cancer Institute's (NCI) Breast Cancer Prevention Trial during the 3-week period after a series of five programs aired *(23)*. According to NCI's Cancer Information Service, the number of Spanish-speaking callers jumped 700% after Dr. Huerta's radio reports began to promote the service's toll-free number *(24)*.

Dr. Huerta's success is especially striking given the often fatalistic way Hispanics view cancer. Much of the success of Dr. Huerta's health communication can be attributed to his under-

standing and incorporation of Hispanic cultural characteristics into message delivery. Feedback from the Hispanic community indicates their desire for health messages to begin with the larger context in which the health message is embedded and move gradually toward the hook or the bottom line—Hispanics, then, fall into the category of a collectivist culture and prefer this style of communication. Dr. Huerta's shows start broad and culminate in the "Big Bang Health Message." Hispanics, moreover, view physicians with authority and respect *(6,25)*. Because Dr. Huerta is Hispanic and a physician, the messages that he disseminates are credible. In other words, because the recipients trust Dr. Huerta as the messenger, his health messages reach the Hispanic community effectively. "Cuidando su Salud" is proof that knowing and respecting audience characteristics is key to effective health communication.

Conclusion: Tailor Health Messages Based on Cultural Beliefs

Knowledge of ethnic differences in risk communication is paramount to the success of a health message *(26)*. Careful attention to culture-specific conceptualizations of a health issue and to appropriate channels for diffusion of health information will help hone the message. When health information such as cancer risks is communicated, messages (and the words used in those messages) must be tailored in a way that will be culturally meaningful and appropriate—this also almost always means communicating in the native language of the cultural group *(3,21)*. Including members of the cultural group as partners in the development and implementation of intervention strategies, moreover, has resulted in significant desirable changes in health risk behavior in several risk-reduction programs *(3)*.

Effective message development, additionally, needs to consider differences among an ethnic group's subgroups. The diversity within subgroups is derived not only from national origin (e.g., Mexican-American, Puerto Rican, Cuban-American, or Central or South American) but also from socioeconomic status, level of education, place of residence (urban or rural), and level of acculturation *(8,9)*. Williams and Flora *(9)*, for example, used the social marketing principle of audience segmentation and disaggregated a Hispanic audience to examine heterogeneous behaviors and lifestyles. Their analysis resulted in six mutually exclusive subgroups, based on self-reported behavioral changes to improve health. Subgroups differed significantly in communication, behavioral, psychological, and demographic dimensions, indicating they may require unique campaign planning strategies.

There is a fine balance between targeting an entire ethnic group with the same broad message and tailoring specific messages to ethnic subgroups or psychobehavioral segments within an ethnic group. Broad health campaigns, such as "Cuidando su Salud," may serve as catalysts for behavioral change, whereas local, more segmented campaigns may be necessary to reinforce a broader health program. Health communications professionals must bear in mind both the uniqueness of an individual and the uniqueness of the cultural group of the individual. Only when these considerations are taken into account can health messages be translated into behavioral change *(21)*.

References

(1) U.S. Department of Commerce, Bureau of Census. Population projections of the United States by age, sex, race, and Hispanic origins: 1990 to 2050. Washington (DC): U.S. Gov Print Off; 1992. p. 25–1095.

(2) Vaughn E. The significance of socioeconomic and ethnic diversity for the risk communication process. Risk Analysis 1995;15:169–80.

(3) Keller CS, Stevens KR. Cultural considerations in promoting wellness. J Cardiovasc Nurs 1997;11:15–25.

(4) Healthy People 2000: National Health Promotion and Disease Prevention Objectives, Full Report, with Commentary. Washington (DC): Department of Health and Human Services; 1990 Report No.: DHHS Publ No. (PHS) 91-50212.

(5) Manning A. Minorities miss out on health news. USA Today, 1999 April 27; 10D.

(6) Marin G. AIDS prevention among Hispanics: needs, risk behaviors, and cultural values. Public Health Rep 1989;104:411–5.

(7) Trickett EJ, Watts RJ, Birman D. Human diversity: perspectives on people in context. San Francisco (CA): Jossey-Bass; 1994.

(8) Johnson EM, Delgado JL. Reaching Hispanics with messages to prevent alcohol and other drug abuse. Public Health Rep 1989;104:588–94.

(9) Williams JE, Flora JA. Health behavior segmentation and campaign planning to reduce cardiovascular disease risk among Hispanics. Health Educ Q 1995;22:36–48.

(10) Crews DE, Bindon JR. Ethnicity as a taxonomic tool in biomedical and biosocial research. Ethn Dis 1991;1:42–9.

(11) McGoldrick M, Pearce JK, Giordano J. Ethnicity and family therapy. New York (NY): Guilford Press; 1982.

(12) Giger JN, Davidhizar RE. Transcultural nursing: assessment and intervention. St. Louis (MO): Mosby-Year Book; 1995.

(13) Triandis HC. Culture and social behavior. New York (NY): McGraw-Hill; 1994.

(14) Palinkas LA, Pierce J, Rosbrook BP, Pickwell S, Johnson M, Bal DG. Cigarette smoking behavior and beliefs of Hispanics in California. Am J Prev Med 1993;9:331–7.

(15) Tortolero-Luna G, Glober GA, Villarreal R, Palos G, Linares A. Cancer research in Hispanic populations in the United States. Monogr Natl Cancer Inst 1995;18:49–56.

(16) Roche RA, Stovall CE, Suarez L, Goldman DA, Wright SA, Mendez MC. Language differences in interpretation of breast cancer health messages. J Cancer Edu 1998;13:226–30.

(17) Morgan C, Park E, Cortes DE. Cancer research in Hispanic populations in the United States. Monogr Natl Cancer Inst 1995;18:57–63.

(18) U.S. Hispanic Market Media. Miami (FL): Strategy Research Corporation; 1994.

(19) 1990 Census of Population and Housing (CPH-L-80). Selected special characteristics: 1990, United States. Washington (DC): U.S. Bureau of the Census; 1991.

(20) Ward B, Bertera EM, Hoge P. Developing and evaluating a Spanish TEL-MED message on breast cancer. J Commun Health 1997;22:127–35.

(21) Locke DC. Increasing multicultural understanding: a comprehensive model. Newbury Park (CA): Sage Publications; 1992.

(22) Shankar S, Huerta EE, Gutierrez-Mohamed M, Crespo CJ. Community involvement in assessing health needs of immigrant Salvadoreans living in the Washington DC area. Ethnicity Dis 1998;8:1–9.

(23) Huerta EE, Weed DL. Cuidando su salud: Spanish-language radio in preventive medicine and public health. Cancer suppl 1998;83:1805–8.

(24) Skolnick AA. Hard to reach Hispanics get health news via physician's radio, TV shows. JAMA 1997;278:269–72.

(25) Marin G. Perceptions by Hispanics of channels and sources of health messages regarding cigarette smoking. Tobacco Control 1996;5:30–6.

(26) Singer M. Confronting the AIDS epidemic among i.v. drug users: does ethnic culture matter? AIDS Educ Prev 1991;3:258–83.

Dealing With Competing and Conflicting Risks in Cancer Communication

Matthew W. Kreuter

Applied research on cancer risk communication is sparse, and even less is known about effective communication under conditions of multiple risks. This paper briefly describes the need and rationale for cancer risk communication, then describes what is known and needs to be known about communication addressing multiple risks. Its focus is on two specific communication issues: 1) comparing different risks and 2) prioritizing between multiple risks. There is considerable unmet need in cancer risk communication for new knowledge and recommendations for best practices. Those professionals choosing to pursue this work can make a significant contribution to the field. [Monogr Natl Cancer Inst 1999;25:27–35]

Two fundamental assumptions underlie the recent emphasis on risk communication in cancer prevention and control.

Assumption 1: Many individuals lack a clear understanding of cancer risks in general and, more specifically, about their own personal cancer risks.

Assumption 2: By better understanding their own cancer risks, individuals will be better able to make informed decisions about screening and about adopting risk-reducing lifestyle changes.

A large body of empirical evidence and theoretical support is available for these assumptions. Much less is known, however, about the relative effectiveness of specific communication strategies designed to help individuals understand their risks. Moreover, what little research has been conducted in cancer risk communication has usually focused on one specific behavioral risk factor (e.g., smoking) or the risk of one specific type of cancer (e.g., breast cancer). The purpose of this paper is to explore what is known and what needs to be known about communicating cancer risk *when an individual has multiple competing risks.* As above, two basic assumptions drive this inquiry.

Assumption 3: Many individuals have multiple cancer risk factors or are nonadherent for multiple types of cancer screening.

Assumption 4: Health communication addressing multiple competing risks requires different considerations or approaches than communication about a single risk.

We will briefly present the evidence supporting assumptions 1–3, then present a more detailed discussion of assumption 4. In this latter section, we will focus on past research and future directions as they relate to two specific issues: 1) comparing different risks and 2) prioritizing between multiple risks.

Assumption 1

Cancer risk information is often misunderstood by the general public, including those at highest risk *(1,2).* Many Americans have misconceptions about causes of cancer and about effective prevention strategies *(3).* For example, as many as half of survey respondents have agreed that "there is not much a person can do to prevent cancer" and that "it seems like everything causes cancer" *(3).* In the case of breast cancer, many women—particularly those who are older, African American, or both—have inaccurate beliefs about the prevention, etiology, diagnosis, and treatment of the disease *(4–11).* In one recent study *(12),* nearly half of urban minority women believed breast cancer almost always led to breast removal, three quarters did not recognize family history as a risk factor, and more than 80% believed that bumping or bruising the breast could lead to breast cancer.

Although an abundance of health information is already available, current offerings may not adequately meet the needs of individuals and their families. For example, more than 1500 consumer-directed health and medical periodicals are listed in the *Standard Periodical Directory (13).* Many of these periodicals address cancer prevention and treatment and are in the public domain. However, information specialists have noted that such materials are often difficult to locate, frequently require "intermediaries" for interpretation, and often lack personal relevance for the individual reader *(14).* Moreover, many readers do not understand technical words used in patient education and cancer prevention materials *(15,16).* One study *(17)* has found significant misconceptions of even fairly common medical terms such as "virus" and "tumor." Although discouraging, these findings are not surprising, given that patient education materials are typically written at a 12th grade reading level *(18–21)* and thus may exceed the capabilities of many members of the general population.

McCallum *(22)* summarized the need for more effective risk communication by saying, "nonexperts need access to information and need to gain knowledge, while technical experts and officials need to learn more about non-experts' interests, values, and concerns." At present, the public may not get enough clear information on which to base difficult, value-laden decisions about which personal health risks are acceptable and which ones are not *(23,24).*

Regarding personal health risks, many studies *(25–35)* have shown that people tend to have inaccurate perceptions of their risks. In most cases, people tend to underestimate their risk, an "optimistic bias" *(29).* However, a few studies *(2,36,37)* suggest that, when estimating personal *cancer* risk, just the opposite is true—people are as likely or more likely to overestimate their risk, a "pessimistic bias." Whether these two risk perception biases are because of cognitive errors such as egocentrism *(29,31),* flawed mental models of information processing *(38),* or simply reflect a psychologically protective coping mechanism *(39),* both biases reflect a distorted understanding of personal risk. People with an optimistic bias about a particular health risk

Correspondence to: Matthew W. Kreuter, Ph.D., M.P.H., Health Communication Research Laboratory, School of Public Health, Saint Louis University, 321 North Spring Ave., St. Louis, MO 63108 (e-mail: kreuter@slu.edu).

© Oxford University Press

(i.e., they perceive their risk to be low, but it is actually high) will have low perceived susceptibility *(26)* and are probably unlikely to make behavioral changes. At the other extreme, some people will have a pessimistic bias about their health risks, believing their risk of some condition is high when it is actually low. People with this type of belief—the so-called "worried well"—may suffer unnecessary concern and anxiety. Preliminary evidence suggests that they may also make more health care visits than those with optimistic or realistic perceptions of personal health risk *(36)*. Ideally, effective risk communication should identify and address both types of bias, with the objective of moving individuals toward a more accurate perception of their own risk. If a person's risk is high, he or she should know it is high. If his or her risk is low, he or she should know that, too.

Assumption 2

Informed Decision Making

More and more individuals are facing difficult decisions about cancer screening, testing, and treatment. Ideally, individuals should understand not only their cancer risks but also the relative harm and benefit associated with each possible course of action that they might choose *(40,41)*. This understanding involves two important steps: 1) knowing which risks are small and which ones are large (i.e., a "quantitative assessment" of risk) and 2) considering more thoroughly the multitude of factors that may influence those risks or result from them (i.e., a "qualitative assessment" of risk) *(41,42)*. Traditionally, discussions of such trade-offs have been lacking in medical encounters between physicians and patients *(43,44)*. Because little is known about how to communicate personal risk information in ways that will facilitate informed decision making, this aspect of risk communication has been identified as a priority for behavioral research in cancer prevention and control *(42,45,46)*.

In the area of cancer screening, a wide range of procedures and tests is now available for early detection of cancer and cancer risk. But, because the risks and benefits of some of these tests are mixed, it is probably more appropriate to promote informed decision making than to indiscriminantly prescribe these tests *(47)*. Supporting this approach, one study *(48)* has shown that nondirective counseling to enhance informed decision making can significantly increase a woman's perceptions of the limitations and risks of BRCA1 gene testing, yet not diminish her interest in being tested.

Of particular interest in this area are yet unanswered questions related to risk communication about genetic testing. For example, although a positive genetic test result can indicate a very high relative risk of getting a particular disease, uncertainty remains about when or if such disease may occur *(49)*. In addition to introducing a serious health threat and uncertainty about the future, genetic testing may also bring on psychological distress related to the new knowledge that one could pass on to his or her offspring an increased risk of certain cancers *(50)*. Several studies have reported that those patients receiving positive results of genetic testing may experience a range of responses, including test-related psychological distress *(51)*, troubling thoughts *(52)*, anxiety, and depressive symptoms *(53)*. Less is known about how best to communicate the results of genetic testing in ways that promote informed decision making or behavioral change but also minimize possible adverse psychological effects. In one study *(54)*, smokers who received biomarker feedback about lung cancer susceptibility as part of a cessation counseling intervention had increased perceived risk, perceived quitting benefits, and fear arousal compared with smokers who did not receive this information but were no more likely to have quit smoking at a 2-month follow-up. Heightening risk perception without adequately addressing specific methods or building self-efficacy for risk reduction is a formula for behavioral inaction *(55)*. With the potential psychological effect of screening also to be considered, the task of effective genetic risk communication is clearly a complex one.

Lifestyle and Behavioral Changes

Most theories *(56–58)* of health-related behavior change identify perceived risk as an important factor in explaining the behavior change process. According to these theories, when individuals perceive their risk of some adverse health outcome to be high, they are more likely to take preventive action to reduce the risk. When perceived risk is accompanied by favorable beliefs about the benefits of that action, the likelihood of change occurring is further enhanced. To illustrate this relationship, in a recent survey *(59)* of Missouri farmers, those who expressed personal concern about skin cancer and felt it was preventable were more likely to use skin protection (sunscreen, hat, long sleeves, or other protective clothing) than farmers who were not personally concerned about it or did not think it was preventable. In practice, the relationship between perceived risk and behavior change can be more complex. For example, perceived susceptibility, a central construct in the health belief model *(56)*, appears to be a stronger predictor of preventive health behaviors than it is of sick role behaviors *(60,61)*, yet, even among preventive behaviors, perceived susceptibility has been shown to be associated with some practices (e.g., mammography) but not others (e.g., breast self-examination) *(62)*.

In addition, although the relationship between perceived risk and behavior change is typically viewed from the standpoint of the former causing the latter, the opposite can also be true *(36)*. For example, once a smoker has quit, his or her perceived susceptibility to tobacco-related illnesses may be reduced. Thus, perceptions may change behavior or vice versa. This dynamic relationship between perceived risk and behavior may also be influenced by the nature of the risk under consideration. For example, people generally underestimate their risk of diseases they perceive as controllable or those that they have direct experience with, and they overestimate their risk of conditions that are less familiar and perceived as beyond their control *(27,29,30,32,34,63)*, yet the specific roles of controllability and familiarity in people's formation of perceived cancer risk are not always clear. For example, how would a woman with a family history of breast cancer (low controllability, high familiarity) be expected to perceive her own risk of breast cancer? Findings have been mixed as to whether this risk factor increases, reduces, or has no effect on perceived vulnerability *(64–66)*. Some women may not consider or understand the possibility of inheriting such a trait, more vigilant others may believe they have reduced their inherited risk by making lifestyle changes, and still others may choose to ignore or deny this information *(12,67)*. Whereas changes in perceived risk alone will seldom be sufficient to bring about changes in cancer prevention behaviors, they can be an important precursor to these changes.

Assumption 3

A person may have multiple cancer risk factors, multiple cancer screening needs, elevated risks for multiple cancers, or some combination of all of these. By using data from an intervention study *(36)* seeking to change inaccurate perceptions of health risks, the researchers examined the frequency of multiple overall health risks among 161 (out of 1131) adult primary care patients who had a higher than average overall cancer risk. In this study, cancer mortality risk was calculated as a composite measure of the 10-year mortality risks for cancers of the mouth, throat, esophagus, stomach, colon, rectum, pancreas, larynx, lung, skin, breast, cervix, uterus, prostate, bladder, and brain. Each individual cancer risk was computed on the basis of known risk factors (e.g., age, family history of cancer, tobacco use, alcohol consumption, menarche, and parity) and protective factors (e.g., hysterectomy and mammography), and a composite risk was derived from these factors. The methodology for this assessment has been described elsewhere *(36,68)*. Among patients whose cancer risk was higher than average, 44% were also at higher than average mortality risk for either heart disease, stroke, or motor vehicle crash. Forty-two percent were at higher than average mortality risk for *two* of these additional health problems, and 4% were at elevated risk for all three. Only 11% of the patients with elevated cancer risk were at average or below-average risk for these other problems.

Because the mortality risk assessed in these data is composed of multiple risk factors (some of which—e.g., smoking—are common to more than one mortality cause), such composite analyses mask the fact that any one person may also have multiple risk *factors*. A person may have different risk factors for multiple diseases or multiple risk factors for the same disease. For example, in data collected from a different intervention trial *(69,70)* among adult primary care patients in southeastern Missouri, many participants had at least one behavioral risk factor for several different types of cancer. Among 101 smokers, more than half (52%) were also physically inactive or ate a high-fat diet. Specifically, 15% of all smokers were also inactive, another 21% ate diets too high in fat, and an additional 17% were inactive *and* ate a diet too high in fat. Nonadherence to cancer screening guidelines is also a type of risk factor. When data from a study of physician gender and screening among primary care patients *(71)* were examined, the prevalence of patients in need of multiple disease screenings was also high. About half (52%) of 2082 patients were in need of either a cholesterol test, mammogram, or Pap smear, based on current screening guidelines; of these, multiple screenings were indicated for more than one third (34%) of the population.

Alternatively, individuals may have multiple risk factors for any single type of cancer. For example, in previously published data from a multicenter screening program for breast cancer *(72)*, a majority of 948 matched control subjects had more than one breast cancer risk factor. Among these 948 women, 3% had none of four risk factors [age at menarche <14 years, first live birth ⩾ 20 years *(3,20)*, more than one previous biopsy, or a mother or sister with breast cancer], and 28% had only one risk factor. However, 60% of women had two risk factors, and 9% had three or more risk factors. Although these latter numbers may be inflated because of the fact that the women participated in a screening program (i.e., they may have had raised consciousness about breast cancer—perhaps because of the presence of one or more risk factors—that led them to screening), the pattern is consistent.

Assumption 4

Effectively communicating cancer risk is challenging even when addressing a single, well-known, and clearly understood risk factor. It is considerably more complex when multiple competing risks are present. This section addresses two specific issues related to communicating multiple risks: 1) comparing different risks and 2) prioritizing between multiple risks.

Comparing Different Risks

Comparing One Risk to Another

When a person faces multiple cancer risks or when multiple screening procedures are indicated, health communicators may present a comparison of these risks to help the individual make well-informed decisions and take the most appropriate action. In such cases, the tendency of health professionals may be to establish priorities on the basis of the relative *magnitude* of each risk. But larger risks will not necessarily be perceived as more serious ones by the public *(73)*. As Maibach and Holtgrave *(74)* have suggested, risk communication should not focus solely on risk probability alone but should also take into account the probability of gain and loss, benefit, and harm associated with a given decision or adoption of some risk-reducing behavior. For example, for a given woman over age 50 years, the attributable cancer risk because of nonadherence with mammography guidelines likely exceeds the risk of nonadherence with Pap testing recommendations. However, for this woman, the fear of finding cancer, losing a breast, feeling less attractive, and possibly losing her partner may be perceived as a substantial loss or harm that could result from getting a mammography. In this case, the health professional's probability-based comparison may be a poor fit for the woman and is more likely to be rejected.

In addition to the magnitude of risk and the relative potential for it to result in benefit or harm, a multitude of other factors are well documented in the research literature that influence individuals' perceptions of personal health risks. These factors should also be carefully considered in any presentation of comparative risk. They include the following:

- **Perceived characteristics of a risk.** Risks that are viewed as more controllable, less likely to have direct or immediate consequences, less likely to lead to catastrophic or devastating outcomes, more fairly distributed in the population, and more familiar to the target population are perceived as more acceptable (i.e., risks they are willing to take) *(22,23,75–78)*.
- **Nature of risk exposure.** Risks that are taken on voluntarily are viewed as more acceptable *(77)*.
- **Actual levels of risk.** Risk probabilities that are very small may be viewed as less credible or less certain *(79)*, and lay people may have an especially difficult time understanding them *(80)*.
- **Understanding probability and uncertainty.** Individuals who are more familiar with the concepts of uncertainty in risk assessment and uncertainty in science may be more likely to understand health risk information *(79)*.
- **Presentation of risk.** Individuals may tend to overestimate the probability of risks presented in relative terms compared with those presented in absolute terms *(81)*.

To address the wide range of factors that might influence a person's perception of his or her own health risks, some researchers *(74,82)* have suggested a balance-sheet approach that considers not only risk probabilities but also the potential for resulting benefit and harm. Additionally, it would likely be helpful to understand how a person perceived a given risk (i.e., which if any of the above-mentioned factors were being used). This type of approach to risk communication would necessarily be highly variable from one individual to the next. In other words, the set of factors influencing person A's perception of a given risk might be very different than the set of factors influencing person B's perception. This strategy is already being used effectively in health behavior research in which computer programs can individually tailor intervention materials that promote a range of cancer-prevention behaviors *(83)*.

The way a health risk or risk-reduction strategy is presented, or framed, can also influence a person's decisions and behavior *(84)*. "Gain" frame messages are those that emphasize attaining a desirable outcome (e.g., "If you get a mammogram, the chances are good that you will find out your breasts are healthy") or avoiding an undesirable one (e.g., "If you get a mammogram, you will decrease your risk of having an undetected breast tumor that could cause cancer") *(85)*. In contrast, "loss" frame messages emphasize undesirable outcomes (e.g., "If you don't get a mammogram, you will increase your risk of having an undetected breast tumor that could cause cancer") or the avoidance of a positive outcome (e.g., "If you don't get a mammogram, you will never find out if your breasts are healthy"). Effects of such message-framing strategies are well documented and appear to vary, based on the type of health behavior being addressed. For example, because screening behaviors tend to be viewed by participants mainly in terms of their undesirable outcomes (i.e., their results inform people if they are sick) *(85)*, loss frame messages should have an advantage over gain frame messages in promoting screening. Data from studies *(86–92)* in the areas of breast cancer detection, cholesterol screening, self-examination for skin cancer, and human immunodeficiency virus testing have all supported this hypothesis. In contrast, because preventive behaviors may be viewed by participants primarily in terms of their benefits (i.e., building or maintaining positive health status and reducing the risk of future problems), gain frame messages should be more effective in promoting these changes. Again, several studies *(86,91,93)* have reported findings consistent with this prediction.

Even within the broad categories of screening and preventive behaviors, effectiveness of different framing strategies may be influenced by individuals' existing beliefs about the particular risk or behavior. For example, one's level of concern about a given health problem and his or her perception of personal risk for that problem may mediate the effects of framing *(89,94)*. As our understanding of framing effects is still evolving, this area appears to be important for future research in risk communication.

The degree of emphasis placed on the *uncertainty* of a given risk can also influence a person's perceptions of the risk. For example, messages can either call attention to the *positive probability* that a risk causes cancer (e.g., "It is probable that *X* causes cancer") or to the lack of complete certainty (e.g., "It is not known for certain whether *X* causes cancer") *(95)*.

An intuitively appealing but as yet untested approach involves the use of qualitative versus quantitative statements of risk probability *(74)*. Considerable research has been conducted linking numeric probabilities to verbal statements expressing probability. The premise behind this research is that people think and talk about the probability of future events every day, but they seldom use the kinds of quantitative estimates (e.g., statistics and numeric proportions) with which most health professionals are accustomed to communicating to the public. For example, the likelihood of event A occurring might be seen as "doubtful" while event B is viewed as "almost certainly" going to happen. If numeric values could be associated with these probabilistic expressions, risk communication might be more effective. Mosteller and Youtz *(96)* reviewed 20 studies of this sort and from them derived a table of 52 probabilistic expressions and the average quantitative probability associated with each. Condensed versions of these findings *(97,98)*, such as that shown in Table 1, have been subsequently suggested. Although this work has also been criticized for its use of relative adjectives and adverbs (e.g., "frequent" and "rarely") without defining an implied referent *(99)*, this potential limitation is less problematic if the expressions are used to compare two or more defined risks, because the referent is then quite explicit. However, epidemiology—presumably a major part of the basis for such probabilistic judgments—is not particularly adept at distinguishing different small risks from one another *(100)*; thus, some such ratings of personal risk could be inherently flawed.

Sometimes the comparison of multiple risks is itself a technique for communicating risk information. That is, its purpose is not so much to promote informed decision making as it is to emphasize the importance of a particular risk over another or to create awareness of some lesser known risk by comparing it with a more familiar risk. This process of this comparative risk assessment (CRA) has been defined as "juxtaposing disparate risks for the purpose of declaring which one is the 'larger' or the 'more important'" *(73)*. Although insufficient empirical evidence is available to recommend this approach as a major intervention strategy, a number of researchers have identified important issues to consider in CRA and communication. According to Persson *(80)*, the chosen referent for risk comparisons must be relevant. For example, comparisons of voluntary risks to involuntary ones should be avoided, as should comparisons of risks that are dissimilar in terms of when and where their consequences may appear. Contrary to this recommendation and similar to ones from others *(101)*, Roth et al. *(102)* have noted that people may actually prefer comparisons in which the health risk of interest is compared with an unrelated but familiar risk (e.g., lightning strikes). From a methodological standpoint,

Table 1. Range of probability and probabilistic statements*

Range of probability	Probabilistic statements
0.00–0.05	Almost never
0.05–0.15	Seldom
0.15–0.25	Infrequent
0.25–0.35	Sometimes
0.35–0.45	Less than an even chance
0.45–0.55	Even chance
0.55–0.65	More often than not
0.65–0.75	Often
0.75–0.85	High probability
0.85–0.95	Very high probability
0.95–1.00	(Virtually) certain

*Source: *see* ref *(97)*.

Fischhoff et al. *(41)* have reported that, although many people have a generally accurate sense as to the relative probability of different risks occurring, perceptions of absolute risk can be greatly influenced by the magnitude of the comparative risk selected. In addition, potentially serious miscalculations can arise when the uncertainty inherent in any two distinct risk estimates considered is not taken into account when the two are being compared *(73)*.

Comparing One Person's Actions to Another's

Weinstein *(31)* suggests that underestimation of risk, or "optimistic bias," is in large part a function of cognitive errors. One error studied by Weinstein is egocentrism—an individual's assumption that other people of similar age, sex, and general situation are not taking the same degree of precautionary action as she is. In a study *(31)* testing the effect of information designed to counteract egocentrism by offering subjects information about the precautionary actions of others, optimistic bias was reduced. An unexpected finding of this study was that subjects in a group instructed to simply describe their own risk factors (with no feedback about others' risk factors) had the greatest level of optimistie bias. It is possible that consideration of one's own risk factors (or lack thereof) actually increases egocentrism and, therefore, optimistic bias.

Most risk communication does not provide feedback regarding the health habits of others. Rather, its focus is on the individual and his or her health habits. Based on the Weinstein *(31)* study, it seems possible that this focus might actually create unintended egocentrism. Recipients of some cancer communications addressing multiple risks may pay particular attention to risk factors they did not have or ones they had never considered previously. For example, a person may focus on the fact that he or she eats lots of vegetables, rather than on the fact that he or she smokes cigarettes. In such a case, risk communication may be inviting egocentrism by presenting risk estimates in terms of absolute rather than relative risk. In other words, although population estimates provide the basis for many risk estimates, risk communications often do not consider this information when formulating risk messages for individuals. In a study of four health risk appraisals (HRAs) administered by Avis et al. *(25)*, feedback from only one HRA was found to have an effect on perceived risk. This HRA specifically informed people that, compared with others, they are at above-average, average, or below-average risk for developing heart disease. This finding would support the hypothesis that optimistic bias is a function of egocentrism, which can, in turn, be counteracted by comparison with others.

Prioritizing Between Multiple Risks

In the presence of a single modifiable behavioral risk, the course of preventive action is fairly clear—modify that behavior. However, addressing multiple risk factors at the same time can burden and confuse individuals, and many will be unable to make changes in multiple risk-related behaviors (e.g., quitting smoking, reducing dietary fat intake, exercising, and reducing alcohol intake) simultaneously *(103–105)*.

In the previously described intervention study seeking to change inaccurate perceptions of health risks *(36)*, patients received feedback indicating that their 10-year mortality risk was higher than average, average, or lower than average for each of four health problems: cancer, heart disease, stroke, and motor vehicle crash. It is interesting that patients who initially overestimated their cancer risk (i.e., a pessimistic bias) and subsequently received feedback explaining that their risk of not only cancer but also heart disease, stroke, and motor vehicle crash were all below average were the most likely to alter their perceptions of personal risk toward a more accurate viewpoint. However, a similar change in risk perception was not found for those who initially underestimated their cancer risk. These patients were no more likely to alter inaccurate perceptions of risk if their feedback indicated four higher-than-average risks or only one higher-than-average risk. In other words, receiving consistent feedback that a person is at lower-than-average risk for multiple different diseases seems to influence perceptions of susceptibility, but receiving consistent feedback that a person is at higher-than-average risk for multiple different diseases does not. This finding may reflect a kind of defensive avoidance reaction when people are faced with threatening information *(39)*.

Given this pattern of risk perception, risk communications and other intervention materials should prioritize among competing risks when multiple risk factors are present. At least five specific criteria could be used to establish such a prioritization: 1) epidemiologic risk, 2) readiness to make behavioral changes, 3) self-efficacy for making behavioral changes, 4) objective difficulty of making behavioral changes, and 5) gateways to behavioral changes. Each is discussed in detail below.

Epidemiologic Risk

Some types of risk communication, like HRA feedback, have traditionally done a good job of implicitly prioritizing between competing risks by quantifying the unique contribution of specific risk factors and risk behaviors to a projected overall risk estimate like "risk age" or an equivalent metric. Some HRA feedback has also taken the next step, describing the extent to which risk age could be reduced by changing various risk factors. Setting risk reduction priorities based solely on an epidemiologic criteria (e.g., attributable risk) would mean that cancer communication would encourage individuals to focus their efforts on whatever factors contributed most to their risk. If successful in getting people to change, this approach would yield the greatest reduction of population risk. However, if users are not interested in changing their "most important" risk factor, they are highly unlikely to do so even with intervention *(106)*.

Readiness to Make Behavioral Changes

As first suggested in the research design of Kreuter and Strecher *(107)*, the focus of supplemental behavior change feedback provided with health risk information can be determined by a person's readiness to change different risk behaviors. Rather than automatically selecting the person's greatest health risk (i.e., epidemiologic, or attributable, risk), this approach would select whichever risk an individual was most interested in changing. This approach should yield the greatest probability of change occurring because the intervention will be helping the person make changes he or she is already considering. For example, the recipient of a risk communication might be a cigarette smoker, eat a high-fat diet, and consume excessive amounts of alcohol. If this person was thinking about reducing his or her dietary fat consumption (i.e., in the contemplation stage for this behavior) but not thinking about quitting or cutting down on his

or her tobacco and alcohol consumption (i.e., a "precontemplator" for these behaviors), his or her risk information might initially focus on helping make the dietary changes he or she wants to make.

Self-Efficacy for Making Behavioral Changes

Efficacy expectations, or "self-efficacy," are beliefs individuals have about their ability to enact a particular behavior *(108)*. They have been found to be strong predictors of a person's success in modifying a range of lifestyle behaviors *(109)*. Behavior-specific self-efficacy could be assessed in conjunction with risk assessments and used as criteria for determining which among multiple risk factors would receive priority attention in risk communication or other educational and behavioral interventions. For example, a person might be very confident that he or she can eat fresh fruit more often, but not at all confident that he or she will be able to give up a lifelong smoking habit. This particular example points out the greatest potential limitation of this approach. Failing to intervene on his or her smoking and instead focusing on fruit and vegetable consumption may seem like shifting deck chairs on the Titanic. However, self-efficacy theory suggests that establishing early successes—even if they are not directly related to other behaviors—can enhance a person's self-efficacy for future prevention activities *(108)*.

Importantly, both the "readiness" and "self-efficacy" prioritization approaches must directly involve individuals in the decision-making process. This approach will probably result in a greater "buy-in" or commitment on the part of the individual *(110)*, which is an important aspect of any intervention involving joint goal setting *(111)*. Moreover, it need not be limited to these two constructs. Tversky *(112)* describes how people use a similar process, "elimination by aspect," to select from alternative risks or behaviors. Given several alternatives to choose from, people are likely to focus on specific desirable (or undesirable) aspects of each alternative, not the alternative itself. For example, if a high level of effort is viewed as undesirable in selecting a behavior change to attempt, some alternatives will take more effort, and they may be eliminated. What is left after multiple rounds of elimination will be the most likely candidate for change.

Objective Difficulty of Making Behavioral Changes

Generally speaking, not all behavior change is equally difficult to accomplish. A smoker can spend 10 minutes one time in his or her life installing a smoke detector and 2 minutes each year checking the battery and will effectively reduce his or her risk of mortality in a house fire. However, behaviors that follow long-time, deeply ingrained patterns and cultural norms (e.g., diet) or behaviors characterized by a strong physiologic dependence (e.g., tobacco use and alcohol consumption) are often more difficult to modify. On the basis of this observation, a person might choose to prioritize between competing risk factors, based on the objective difficulty of changing those behaviors.

Gateways to Behavioral Changes

For many individuals, different health-related conditions and behaviors are closely linked to one another. For example, engaging in physical activity is associated with eating a healthier diet and with various psychological outcomes *(113,114)*, and smoking is associated with a poorer diet, injuries, and alcohol consumption *(115–117)*. The presence of certain health problems can also make it more difficult, if not impossible, for a person to modify some health behaviors. For example, certain types of physical activity will be limited by low-back problems, and sustained cessation is associated (negatively) with depression *(118)*. Left unchecked, depressive symptoms would likely be a substantial barrier to a person quitting smoking or, for that matter, to making any serious attempt to change other health-related behaviors *(119,120)*. But smoking cessation programs that include a mood management intervention appear to be especially effective among depressed smokers *(121)*. Thus, it would seem important to address these "gateway" problems so that other behavioral intervention attempts will stand a better chance to succeed. It may also be possible to better facilitate behavioral changes by introducing new "gateway behaviors," such as physical activity, that if adopted may increase the likelihood that other lifestyle changes are also made *(114)*.

Combinations of Criteria

Naturally, these criteria are not mutually exclusive. Advanced computer technologies make it possible to consider multiple, or even all of these, criteria in a single prioritization algorithm. From a behavioral science standpoint, the key evolutionary step this prioritization offers is movement away from solely risk-based criteria to a more systematic consideration of users' interests, beliefs, and general life circumstance.

Summary

Applied research on cancer risk communication is sparse, and even less is known about effective communication under conditions of multiple risks, yet risk communication-based interventions would seem to be a critical part of strategies to promote adoption of healthy lifestyle behaviors and informed decision making. This paper sought to raise issues for consideration in developing and testing such cancer risk communication and to provide some initial suggestions for guiding that work. There is considerable unmet need in this area for new knowledge and recommendations for best practices. Those professionals choosing to pursue this line of work can make a significant contribution to cancer risk communication.

References

(1) Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995;87:286–92.

(2) Black WC, Nease RF Jr, Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst 1995;87:720–31.

(3) Cancer Prevention Awareness Survey Technical Report. Bethesda (MD): National Cancer Institute Office of Cancer Communications; 1987; Report No.: 87-2907.

(4) Burack RC, Liang J. The acceptance and completion of mammography by older black women. Am J Public Health 1989;79:721–26.

(5) Lacey LP, Phillips CW, Ansell D, Whitman S, Ebie N, Chen E. An urban community-based cancer prevention screening and health education intervention in Chicago. Public Health Rep 1989;104:536–41.

(6) Darrow SL, Schoenfeld ER, Cummings KM, Wilkes E, Madoff S. Women's knowledge and beliefs about breast cancer risk factors, symptoms, detection methods, and treatments. J Cancer Educ 1987;2:165–76.

(7) Taplin SH, Montano DE. Attitudes, age, and participation in mammographic screening: a prospective analysis. J Am Board Fam Pract 1993; 6:13–23.

(8) Fuller SM, McDermott RJ, Roetzheim RG, Marty PJ. Breast cancer beliefs of women participating in a television-promoted mammography screening project. Public Health Rep 1992;107:682–90.
(9) Price JH, Desmond SM, Slenker S, Smith D, Stewart PW. Urban black women's perceptions of breast cancer and mammography. J Community Health 1992;17:191–204.
(10) Miller AM, Champion VL. Attitudes about breast cancer and mammography: racial, income, and educational differences. Women Health 1997; 26:41–63.
(11) Dolan NC, Lee AM, McDermott MM. Age-related differences in breast carcinoma knowledge, beliefs, and perceived risk among women visiting an academic general medicine practice. Cancer 1997;80:413–20.
(12) Skinner CS, Arfken CL, Sykes RK. Knowledge, perceptions, and mammography stage of adoption among older urban women. Am J Prev Med 1998;14:54–63.
(13) Striplin D, editor. The standard periodical directory. 17th ed. New York (NY): Oxbridge Communications; 1994.
(14) Deering MJ, Harris J. Consumer health information demand and delivery: implications for libraries. Bull Med Library Assoc 1996;84:209–16.
(15) Reid JC, Kardash CM, Robinson RD, Scholes R. Comprehension in patient literature: the importance of text and reader characteristics. Health Commun 1994;6:327–35.
(16) Jackson A, Sears S. Implications of an Africentric worldview in reducing stress for African American women. J Couns Dev 1992;71:184–90.
(17) Spiro D, Heidrich F. Lay understanding of medical technology. J Fam Practice 1983;17:277–9.
(18) Davis TC, Crouch MA, Wills G, Miller S, Abdehou DM. The gap between patient reading comprehension and the readability of patient education materials. J Fam Pract 1990;31:533–8.
(19) Meade CD, Diekmann J, Thornhill DG. Readability of American Cancer Society patient education literature. Oncol Nurs Forum 1992;19: 51–5.
(20) Reid JC, Riche J. The readability of Arthritis Foundation pamphlets. Paper presented at the Arthritis Health Professions Association; 1990 May; Chicago (IL).
(21) Butow P, Brindle E, McConnell D, Boakes R, Tattersall M. Information booklets about cancer: factors influencing patient satisfaction and utilization. Patient Educ Couns 1998;33:129–41.
(22) McCallum DB. Risk communication: A tool for behavior change. NIDA Res Monogr 1995;155:65–89.
(23) Rowan F. The high stakes of risk communication. Prev Med 1996;25: 26–9.
(24) Ansell J, Wharton F, editors. Risk: analysis, assessment, management. New York (NY): Wiley; 1992.
(25) Avis NE, Smith KW, McKinlay JB. Accuracy of perceptions of heart attack risk: what influences perceptions and can they be changed? Am J Public Health 1989;79:1608–12.
(26) Becker DM, Levine DM. Risk perception, knowledge, and lifestyles in siblings of people with premature coronary disease. Am J Prev Med 1987;3:45–50.
(27) Kulik JA, Mahler HI. Health status, perceptions of risk, and prevention interest for health and nonhealth problems. Health Psychol 1987;6:15–27.
(28) Lee C. Perceptions of immunity to disease in adult smokers. J Behav Med 1989;12:267–77.
(29) Weinstein ND. Unrealistic optimism about future life events. J Pers Soc Psychol 1980;39:806–20.
(30) Weinstein ND. Unrealistic optimism about susceptibility to health problems. J Behav Med 1982;5:441–60.
(31) Weinstein ND. Reducing unrealistic optimism about illness susceptibility. Health Psychol 1983;2:11–20.
(32) Weinstein ND. Why it won't happen to me: perceptions of risk factors susceptibility. Health Psychol 1984;3:431–57.
(33) Weinstein ND. Unrealistic optimism about susceptibility to health problems: conclusions from a community-wide sample. J Behav Med 1987; 10:481–500.
(34) DeJoy DM. The optimism bias and traffic accident risk perception. Accid Anal Prev 1989;21:333–40.
(35) Svenson O. Risks of road transportation in a psychological perspective. Accid Anal Prev; 1978;10:267–80.
(36) Kreuter MW, Strecher VJ. Changing inaccurate perceptions of health risk: results from a randomized trial. Health Psychol 1995;14:56–63.
(37) Skinner CS, Kreuter MW, Kobrin S, Strecher VJ. Perceived and actual breast cancer risk. Optimistic and pessimistic biases. J Health Psychol 1998;3:181–93.
(38) Atman CJ, Bostrom A, Fischhoff B, Morgan MG. Designing risk communications: completing and correcting mental models of hazardous processes, Part 1. Risk Anal 1994;14:779–88.
(39) Taylor SE. Adjusting to threatening events: a theory of cognitive adaptation. Am Psychol 1983 Nov:1161–73.
(40) Unger JB. Mammography screening for women in their 40s: facilitating truly informed decision making. Am J Public Health 1998;88:840–1.
(41) Fischhoff B, Bostrom A, Quadrel MJ. Risk perception and communication. Annu Rev Public Health 1993;14:183–203.
(42) Lerman C, Rimer B, Glynn T. Priorities in behavioral research in cancer prevention and control. Prev Med 1997;26 (5 Pt 2):S3-9.
(43) Katz J. The silent world of doctor and patient. New York (NY): Free Press; 1984.
(44) Braddock CH 3rd, Fihn SD, Levinson W, Jonsen AR, Pearlman RA. How doctors and patients discuss routine clinical decisions. Informed decision making in the outpatient setting. J Gen Intern Med 1997;12: 339–45.
(45) National Research Council. Improving risk communication. Washington (DC): National Academy Press; 1989.
(46) Lerman C. Translational behavioral research in cancer genetics. Prev Med 1997;26:S65–9.
(47) Ernster VL. Mammography screening for women aged 40 through 49—a guidelines saga and a clarion call for informed decision making. Am J Public Health 1997;87:1103–6.
(48) Lerman C, Biesecker B, Benkendorf JL, Kerner J, Gomez-Caminero A, Hughes C, et al. Controlled trial of pretest education approaches to enhance informed decision-making for BRCA1 gene testing. J Natl Cancer Inst 1997;89:148–57.
(49) Lerman C. Psychological aspects of genetic testing: introduction to the special issue. Health Psychol 1997;16:3–7.
(50) Lerman C, Croyle RT. Emotional and behavioral responses to genetic testing for susceptibility to cancer. Oncology 1996;10:191–5, 9.
(51) Croyle RT, Smith KR, Botkin JR, Baty B, Nash J. Psychological responses to BRCA1 mutation testing: preliminary findings. Health Psychol 1997;16:63–72.
(52) Marteau TM, Dundas R, Axworthy D. Long-term cognitive and emotional impact of genetic testing for carriers of cystic fibrosis: the effects of test result and gender. Health Psychol 1997;16:51–62.
(53) Codori AM, Petersen GM, Boyd PA, Brandt J, Giardiello FM. Genetic testing for cancer in children. Short-term psychological effect. Arch Pediatr Adolesc Med 1996;150:1131–8.
(54) Lerman C, Gold K, Audrain J, Lin TH, Boyd NR, Orleans CT, et al. Incorporating biomarkers of exposure and genetic susceptibility into smoking cessation treatment: effects on smoking-related cognitions, emotions, and behavior change. Health Psychol 1997;16:87–99.
(55) Strecher VJ, Becker MH, Kirscht JP, Eraker SA, Graham-Tomasi RP. Psychosocial aspects of changes in cigarette-smoking behavior. Patient Educ Couns 1985;7:249–62
(56) Becker MH. The health belief model and personal health behavior. Health Educ Monogr 1974;2:324–473.
(57) Fishbein M, Azjen I. Beliefs, attitudes, intention, and behavior: an introduction to theory and research. Reading (MA): Addison-Wesley; 1975.
(58) Weinstein ND. The precaution adoption process. Health Psychol 1988;7: 355–86.
(59) Wadud SE, Krueter MW, Clarkson S. Risk perception, beliefs about prevention, and preventive behaviors of farmers. J Agr Safety Health 1998; 4:15–24.
(60) Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;1:1–47.
(61) Curry SJ, Emmons KM. Theoretical models for predicting and improving compliance with breast cancer screening. Ann Behav Med 1994;16: 302–16.
(62) Savage SA, Clarke VA. Factors associated with screening mammography and breast self-examination intentions. Health Educ Res 1996;11: 409–21.

(63) Weinstein ND. Effects of personal experience on self-protective behavior. Psychol Bull 1989;105:31–50.
(64) Evans DG, Burnell LD, Hopwood P, Howell A. Perception of risk in women with a family history of breast cancer. Br J Cancer 1993;67: 612–4.
(65) Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med 1991;114:657–61.
(66) Ponder M, Lee J, Green J, Richards M. Family history and perceived vulnerability to some common diseases: a study of young people and their parents. J Med Genet 1996;33:485–92.
(67) Chavez LR, Hubbell FA, McMullin JM, Martinez RG, Mishra SI. Understanding knowledge and attitudes about breast cancer. Arch Fam Med 1995;4:145–52.
(68) Amler R, Moriarty D, Hutchins E, editors. Healthier people. Decatur (GA): Carter Center of Emory University Health Risk Appraisal Program; 1989.
(69) Kreuter MW, Brennan LK, Scharff DS, Lukwago SN. Do nutrition label readers eat healthier diets? Behavioral correlates of adults' use of food labels. Am J Prev Med 1997;13:277–83.
(70) Kreuter MW, Scharff DP, Brennan LK, Lukwago SN. Physician recommendations for diet and physical activity: which patients get advised to change? Prev Med 1997;26:825–33.
(71) Kreuter MW, Strecher VJ, Harris R, Kobrin SC, Skinner CS. Are patients of women physicians screened more aggressively? A prospective study of physician gender and screening. J Gen Intern Med 1995;10:119–25.
(72) Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case–control data. Am J Epidemiol 1985;122:904–14.
(73) Finkel AM. Toward less misleading comparisons of uncertain risks: the example of aflatoxin and alar. Environ Health Perspect 1995;103: 376–85.
(74) Maibach E, Holtgrave DR. Advances in public health communication. Annu Rev Public Health 1995;16:219–38.
(75) Tversky A, Kahneman D. Judgement under uncertainty: heuristics and biases. Science 1974;185:1124–31.
(76) Vlek C, Stallen PJ. Rational and personal aspects of risk. Acta Psychologica 1980;45:273–300.
(77) Slovic P, Fischhoff B, Lichtenstein S. Facts versus fears: understanding perceived risk. In: Kahneman D, Slovic P, Tversky A, editors. Judgements under uncertainty: heuristics and biases. New York (NY): Cambridge University Press; 1982, p. 463–89.
(78) Klein WM, Kunda Z. Exaggerated self-assessments and the preference for controllable risks. Organizational behavior and human decision processes. J Pers Soc Psychol 1994;59:410–27.
(79) Johnson BB, Slovic P. Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. Risk Anal 1995; 15:485–94.
(80) Persson L. On communication and comparison of radiation risks. Health Phys 1993;64:528–30.
(81) Malenka DJ, Baron JA, Johabsen S, Wahrenberger JW, Ross JM. The framing effect of relative and absolute risk. J Gen Intern Med 1993;8: 543–8.
(82) Holtgrave DR, Weber EU. Dimensions of risk perception for financial and health risks. Risk Anal 1993;13:553–8.
(83) Skinner CS, Campbell MK, Rimer BK, Curry J, Prochaska JO. How effective is tailored print communication? Ann Behav Med. In press 1999.
(84) Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981;211:453–8.
(85) Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997;121:3–19.
(86) Rothman AJ, Salovey P, Antone C, Keough K, Martin CD. The influence of message framing on intentions to perform health behaviors. J Exp Soc Psychol 1993;29:408–33.
(87) Banks SM, Salovey P, Greener S, Rothman AJ, Moyer A, Beauvais J, et al. The effects of message framing on mammography utilization. Health Psychol 1995;14:178–84.
(88) Meyerowitz BE, Chaiken S. The effect of message framing on breast self-examination attitudes, intentions, and behavior. J Pers Soc Psychol 1987;52:500–10.
(89) Meyerowitz BE, Wilson DK, Chaiken S. Loss-framed messages increase breast self-examination for women who perceive risk. Paper presented at the annual convention of the American Psychology Society; 1991; Washington (DC).
(90) Maheswaran D, Meyers-Levy J. The influence of message framing and issue involvement. J Marketing Res 1990;27:361–7.
(91) Block LG, Keller PA. When to accentuate the negative: the effects of perceived efficacy and message framing on intentions to perform a health-related behavior. J Marketing Res 1995;32:192–203.
(92) Kalichman SC, Coley B. Context framing to enhance HIV-antibody-testing messages targeted to African American women. Health Psychol 1995;14:247–54.
(93) Robberson MR, Rogers RW. Beyond fear appeals: negative persuasive appeals to health and self-esteem. J Appl Soc Psychol 1988;13:277–87.
(94) Rothman AJ, Klein WM, Weinstein ND. Absolute and relative biases in estimations of personal risk. J Appl Soc Psychol 1996;26: 1213–36.
(95) Spedden SE, Ryan PB. Probabilistic connotations of carcinogen hazard classifications: analysis of survey data for anchoring effects. Risk Anal 1992;12:535–41.
(96) Mosteller F, Youtz C. Quantifying probabilistic expressions. Stat Sci 1990;5:2–34.
(97) Kadane JB. Comment: quantifying probabilistic expressions. Stat Sci 1990;5:18–20.
(98) Mosteller F, Youtz C. Rejoinder: quantifying probabilistic expressions. Stat Sci 1990;5:32–4.
(99) Clark HH. Comment: quantifying probabilistic expressions. Stat Sci 1990; 5:12–6.
(100) Taubes G. Epidemiology faces its limits. Science 1995;269:164–9.
(101) Slovic P, Kraus N, Covello V. Comment: what should we know about making risk comparisons? Risk Anal 1990;10:389–92.
(102) Roth E, Morgan MG, Fischhoff B, Lave L, Bostrom A. What do we know about making risk comparisons? Risk Anal 1990;10:375–92.
(103) Currie BF, Beasley JW. Health promotion in the medical encounter. In: Taylor RB, Ureda JW, editors. Health promotion: principles and clinical applications. Norwalk (CT): Appleton-Century-Crofts, 1982, p. 143–60.
(104) Fraser GE. Preventive cardiology. New York (NY): Oxford University Press; 1986. p. 319–45.
(105) Haynes RB. Strategies for improving compliance: a methodological analysis and review. In: Sackett DL, Haynes RB, editors. Compliance with therapeutic regimes. Baltimore (MD): John Hopkins University Press; 1976.
(106) Prochaska JO, Velicer WF, DiClemente CC, Fava J. Measuring processes of change: applications to the cessation of smoking. J Consult Clin Psychol 1988;56:520–8.
(107) Kreuter MW, Strecher VJ. Do tailored behavior change messages enhance the effectiveness of health risk appraisal? Results from a randomized trial. Health Educ Res 1996;11:97–105.
(108) Bandura A. Toward a unifying theory of behavior change. Psychol Rev 1977;84:191–215.
(109) Strecher VJ, DeVellis BM, Becker MH, Rosenstock IM. The role of self-efficacy in achieving health behavior change. Health Educ Q 1986; 13:73–92.
(110) Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. Ann of Behav Med. In press 1999.
(111) Strecher VJ, Seijts GH, Kok GJ, Latham GP, Glasgow R, DeVellis B, et al. Goal setting as a strategy for health behavior change. Health Educ Q 1995;22:190–200.
(112) Tversky A. 1972. Elimination by aspects: a theory of choice. Psychol Rev 1972;79:281–99.
(113) Johnson MF, Nichols JF, Sallis JF, Calfas KJ, Hovell MF. Interrelationships between physical activity and other health behaviors among university men and women. Prev Med 1998;27:536–44.
(114) DiLorenzo TM, Bargman EP, Stucky-Ropp R, Brassington GS, Frensch PA, LaFontaine T. Long-term effects of aerobic exercise on psychological outcomes. Prev Med 1999;28:75–85.
(115) Osler M. The food intake of smokers and nonsmokers: the role of partner's smoking behavior. Prev Med 1998;27:438–43.
(116) Leistikow BN, Shipley MJ. Might stopping smoking reduce injury death

risks? A meta-analysis of randomized, controlled trials. Prev Med 1999; 28:255–9.

(117) U.S. Department of Health and Human Services. The health benefits of smoking cessation. Rockville (MD): U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, Office on Smoking and Health; 1990. DHHS Publication No. (CDC)90-8416.

(118) Glassman AH, Helzer JE, Covey LS, Cottler LB, Stetner F, Tipp JE, et al. Smoking, smoking cessation, and major depression. JAMA 1990;264: 1546–9.

(119) Pietromonaco PR, Markus H. The nature of negative thoughts in depression. J Pers Soc Psychol 1985;48:799–807.

(120) Alloy LB, Ahrens AH. Depression and pessimism for the future: biased use of statistically relevant information in predictions for self versus others. J Pers Soc Psychol 1987;52:366–78.

(121) Hall SM, Munoz R, Reus V. Depression and smoking treatment: a clinical trial of an affect regulation treatment. Paper presented at the fifty-third annual scientific meeting of the Committee on Problems of Drug Dependence; 1991; Palm Beach (FL).

New Directions for Risk Communication Research: a Discussion With Additional Suggestions

Alfred C. Marcus

The papers by Huerta and Macario and Kreuter share the theme of suggesting new directions for risk communication research in cancer prevention and control. Huerta and Macario remind us once again that sociocultural factors must be considered when conducting risk communication research on underserved populations. Of special note is their recommendation to target the family, which could introduce a compelling new chapter in risk communication research in cancer prevention and control. In contrast, Kreuter challenges us to consider multiple cancer risks and risk-reducing behaviors in our research and provides a provocative framework for achieving this goal. Given this common theme and the need to position specific recommendations within the larger context of other competing research questions, this paper also highlights several additional recommendations for future research. These recommendations include the following: more research on risk presentation; establishing guidelines for measuring risk; additional research testing strategies to de-bias optimistic and pessimistic perceptions of risk and evaluating risk communication as a strategy for behavior change; more research investigating the sociology of risk communication, with a special emphasis on the family as the unit of investigation; and, finally, more research that specifically targets underserved populations in diverse community settings. [Monogr National Cancer Inst 1999;25:35–42]

I am delighted to have this opportunity to respond to the papers by Huerta and Macario *(1)* and Kreuter *(2)*. Both papers share the common theme of suggesting new directions for research in risk communication. The main thesis of Huerta and Macario is that more research is needed that specifically targets underserved and minority populations, while Kreuter proposes that future research consider the challenge of targeting multiple risk factors. Although both papers approach this general theme from different perspectives, each has substantial merit as we contemplate the future course of risk communication research in cancer prevention and control.

"Communicating Health Risk to Ethnic Groups: Reaching Hispanics as a Case Study"—Elmer E. Huerta and Everly Macario

Huerta and Macario remind us once again that sociocultural factors must be taken into account when designing and implementing risk communication interventions for racial/ethnic minorities. This paper is especially timely, given that much of the risk communication research conducted to date has focused on the majority population. Thus, as this field of research expands to include other populations, the factors identified by Huerta and Macario will need to be given serious consideration. For example, among the more intriguing factors identified by Huerta and Macario was the distinction between "collectivistic" versus "individualistic" cultures. Individuals from a collectivistic culture may respond more favorably to messages that first provide the context for the message, building methodically, as noted by Huerta and Macario, to the main take-home message. In contrast, individuals from individualistic cultures may prefer to hear the best or strongest arguments first, as a way to generate desire or interest to hear the entire argument. Although this distinction appears noteworthy, whether it actually merits meaningful differences in message structure is a legitimate question that could, and perhaps should be, answered through research.

The complexity and challenges of designing culturally competent risk messages is further illustrated by Huerta and Macario with respect to designing health messages for Hispanics. According to Huerta and Macario, such messages should consider several core values within "Hispanic" culture, including *familialism, collectivism, simpatia, personalismo, respeto,* and *power distance.* I was especially struck by their recommendation to consider the family when designing risk communication interventions, which is a recommendation that would seem to transcend cultural boundaries. This recommendation reminds all of

Correspondence to: Alfred C. Marcus, Ph.D., AMC Cancer Research Center, 1600 Pierce St., Denver, CO 80214 (e-mail: marcusa@amc.org).

© Oxford University Press

us that preventive health behaviors are nurtured and sustained within the context of one's larger social environment. Although this recommendation to target the family is appealing, it also presents at least two fundamental challenges.

Broadly speaking, the first challenge is one of conceptualization. What types of information should be collected from families and from whom? Does a family focus simply translate into collecting health risk information from individual family members? I suspect not. For some risk-reducing behaviors (e.g., diet and nutrition), the preferred unit of intervention may in fact be the family, while others will require a focus on the individual (e.g., cancer screening behavior). However, even in those situations where the individual remains as the unit of intervention, the larger context of the family can play a vital role in risk communication, especially with respect to setting priorities for health behavior and sustaining these behaviors over time. It would seem, for example, that framing messages in terms of reducing family risk from cancer would be beneficial in promoting behavior change, especially among populations where protection of the family is emphasized (e.g., eating more fruits and vegetables may help protect your family from cancer). Similarly, risk messages that target specific family members may take on added significance if they can be framed in terms of protecting family functioning (e.g., if you are screened for colorectal cancer on a regular basis, it may help you prevent colorectal cancer from developing and thus protect your family from the consequences of this disease). It might even be possible, under certain circumstances (e.g., with appropriate informed consent), to share individual risk information with other family members, which might then mobilize the protective function of the family.

There are undoubtedly other ways to conceptualize risk messages within the context of the family. However, the three examples cited above serve to illustrate the potential value of this key recommendation. Families are more than the sum of their individual members. If we can harness the sociology of the family in our risk communication interventions, a new and exciting chapter in risk communication research could emerge.

In addition to the above, patterns of family communication and decision making, both of which have a definitive sociocultural context, should be of great interest to those who embrace the family orientation recommended by Huerta and Macario. Indeed, one could argue that before we can effectively target the family, we first need to improve our basic understanding of how the family might impact risk communication. To what extent will individual risk information be filtered through the prism of family dynamics? How will such information be reinforced, distorted, or otherwise modified by other family members? Under what conditions should risk communication programs explicitly take these family dynamics into account?

Precisely the opposite question should also attract our attention in risk communication research. Thus, in addition to asking what impact the family might have on individual risk communication, we could also ask what impact this information will have on the family. As noted recently by Bottorff et al. *(3)* in their discussion of genetic testing for inherited cancers, "Communication of cancer risk can also have serious implications for family relationships... however, nearly all of the empirical literature concerning the impact of testing solely addresses the individual (pg. 73)." This potential impact on the family might be expected for risk factors that are either inherited or result from common exposures. Unfortunately, the diffusion and interpretation of risk information within the family will not always be tempered by scientific judgment and thus could extend to other situations as well. This possibility accentuates the need to consider these potential consequences on the family as a whole, including unwarranted perceptions of risk that might cascade to other family members. Perhaps the greatest sin would be to ignore these potential consequences altogether, which highlights the need for more research to document and assess the diffusion of individual risk information and its potential impact within the family.

A second major challenge attending a focus on the family involves logistical issues. As noted above, one key decision is whom to target in the family for collecting health risk information. Another key logistical question is how to gain access and collect such information from families. For some risk behaviors, individuals might be targeted, others may require targeting the male or female head of household as a spokesperson for the family, while still others may require targeting the family unit rather than individual family members. If the risk behaviors have a normative component, confidentiality of data must be considered, especially if risk profiles are being collected from adolescents (e.g., exposure to tobacco). Issues of confidentiality become even more pronounced when operationalizing procedures for individualized feedback to families and individual family members.

Taken together, these conceptual and logistical challenges highlight the need to assemble multidisciplinary research teams that have expertise not only in risk communication, tailored message technology, behavioral science, and health education but also in medical anthropology and the sociology of the family. It also seems self-evident that we need more minority investigators involved in the research agenda suggested by Huerta and Macario. My own sense is that the focus on the family represents exceptionally fertile ground for future research. However, so little is known in this regard that a substantial amount of pilot research will be needed, perhaps funded by the National Cancer Institute (NCI) by the use of program announcements or request for applications (RFAs) that specifically encourage phase I and phase II research.

Also highlighted in the paper by Huerta and Macario was the importance of considering appropriate channels for conveying risk information to subjects. There is a wide variety of such channels to choose from, including electronic media, print material, telephone counseling, lay health advisor interventions, and other interpersonal education and counseling interventions. Their description of the use of radio was especially intriguing and could serve as a model for other investigators who face the challenge of reaching underserved populations. In addition, I would like to briefly comment on the use of computer systems to tailor risk information and behavioral recommendations to subjects *(4–7)*.

This new technology clearly has great potential to make risk communication "self relevant" to subjects and perhaps more efficacious in promoting behavior change. However, much of the research conducted to date has tested this new technology within the context of the traditional health care system *(4,5,8–17)*. Although such research needs to receive continual support, other venues should also be explored, especially as a vehicle for reaching the populations emphasized by Huerta and Macario. For example, tailored message interventions delivered in a hospital or clinic setting may not reach populations that underutilize

the traditional health care system. Moreover, placing computers in waiting rooms can be expensive, especially in public-supported hospitals and clinics, patient flow may not allow all patients to utilize this technology even if it were made available, and computer phobia could be a deterrent to their utilization. At the very least, computer-tailored interventions will need to be made available in other locations in the community by use of kiosks and other venues *(6)*.

One promising avenue for future research would involve integrating this new technology within existing models of lay health advisor (LHA) interventions *(18,19)*. Thus far, such interventions have relied mainly on personal encounters to convey health-education messages to clients. However, these personal encounters could also be used to generate tailored follow-up print material, including tailored risk communication *(20–22)*. This marriage between LHA interventions and computer-tailored risk communication would seem to represent another profitable direction for future research and may be especially attractive to the emphasis on families suggested by Huerta and Macario.

Another example of moving this technology beyond the health care setting is illustrated by the research we are conducting in partnership with the NCI-funded Cancer Information Service (CIS) *(23)*. Thus, at the end of usual service, a brief baseline telephone assessment is conducted of CIS callers related to fruit and vegetable consumption, colorectal screening, or smoking cessation. This baseline assessment will then be used to generate follow-up tailored print communication, including tailored risk communication.

This same strategy could be applied to an almost infinite variety of community programs and agencies. What would be required is a commitment from program administrators to extend client services to include risk assessment, to provide appropriate time during client encounters to allow for these assessments to occur, and to facilitate feedback of tailored risk information to clients. Obvious candidates for such programs include work-site wellness programs, schools, nutrition programs, churches, etc. *(20–22,24–28)*. Clearly, this focus on integrating computerized risk assessments and tailored risk communication within existing community-based organizations should be given high priority in future research and would represent, at least in my view, a worthwhile concept for phase III–IV research funded by the NCI through program announcements or the RFA mechanism.

"Dealing With Competing and Conflicting Risks in Cancer Communication"—Matthew W. Kreuter

Kreuter addresses a critically important and largely unexplored issue in risk communication: how to deal with competing and conflicting risks in cancer communication. This is a daunting task, given that we are still struggling with de-biasing optimistic perceptions of risk with respect to a single precautionary or risk-reducing behavior *(29)*. Nonetheless, Kreuter makes a convincing case for addressing this issue in future research by introducing four key assumptions:

Assumption 1: Many individuals lack a clear understanding of cancer risks in general, and more specifically, about their own personal cancer risks.

Assumption 2: By better understanding their own cancer risks, individuals will be better able to make informed decisions about screening and about adopting risk-reducing lifestyle changes.

Assumption 3: Many individuals have multiple cancer risk factors, or are nonadherent for multiple types of cancer screening.

Assumption 4: Health communication addressing multiple competing risks requires different considerations or approaches than communication about a single risk.

Kreuter provides substantial evidence supporting the first three assumptions, all of which lead logically to the fourth assumption, which provides the context for Kreuter's main thesis. As risk communicators and behavioral scientists in cancer prevention and control, we need to move our discipline forward by designing, implementing, and evaluating interventions that will provide information on competing risks, prioritize these competing risks, and, by extension, prioritize risk-reducing behaviors for individuals. Thus, Kreuter recommends that we compare competing risks for subjects based on statistical risk probabilities, probabilities of gains/losses (perhaps using a balance sheet to present this information), perceived controllability of competing risks, familiarity with risk and its consequences, voluntary versus involuntary nature of the risk exposure, and actual risk levels. Ultimately, as suggested by Kreuter, "Risk communications and other intervention materials should prioritize among competing risks when multiple risk factors exist," based on such factors as epidemiologic risk, readiness to change, self-efficacy, objective difficulty of behavioral change, and selected "gateways" to behavior change.

This agenda for designing and testing a competing risk intervention is both visionary and, it seems to me, enormously challenging. Two potential obstacles that might limit our ability to implement Kreuter's research agenda are briefly highlighted below.

1) *Will the underlying science in cancer prevention and control, at this point in time, support the model proposed by Kreuter*? This model assumes sufficient precision to differentiate among individual risk factors, and, by extension, specific risk-reducing behaviors, especially in terms of their relative contribution in reducing the individual's overall risk from cancer. However, if there is one defining challenge in risk communication, it involves communicating risk under conditions of uncertainty. This is a profound and pervasive challenge encompassing virtually all behavioral risk factors in cancer control, exclusive perhaps of cigarette smoking, which should be at the top of everyone's list. For example, how much risk is associated with a poor diet, and what elements of a poor diet should be given the highest priority? How much reduction in risk is associated by increasing fruit and vegetable intake from three servings per day to five or from two servings per day to five? Where do we rank sun protection or clinical examinations for skin cancer relative to other precautionary behaviors in cancer prevention and control? With respect to specific cancer screening tests, which of several different guidelines will be used to establish priorities for reducing cancer risk (e.g., ACS, NCI, U.S. Preventive Services Task Force, the subject's managed care organization)? When providing individualized feedback to subjects regarding priorities for risk reduction, how does one modulate specific recommendations based on age, family history, self-efficacy, readiness to change, etc.? How does one determine "objective difficulty" in making specific recommendations for behavior change? Again, the overriding concern is that the underlying science is not suf-

ficiently developed to answer these or a myriad of related questions that would appear to be fundamental to the model proposed by Kreuter.

2) *How exportable is the approach recommended by Kreuter?* At a minimum, this approach would involve assessing multiple cancer risks and risk-reducing behaviors, providing information to subjects regarding these competing risks and recommending priorities for risk-reducing behaviors. However, it would also seem advisable to convey to subjects the underlying rationale for the specific, individualized set of priorities that are being recommended, which, according to Kreuter's model, would involve some combination of additional factors, including self-efficacy, readiness to change, objective difficulty in making changes to specific behaviors, etc. Moreover, once these priorities are established and their rationale is explained to the subjects, it would seem prudent to introduce selected behavior change strategies to promote the high priority behaviors that are being recommended. From a program implementation and management perspective, this approach would appear to be quite challenging, especially if the goal is to implement a low-intensity public health intervention to reduce cancer risk within defined populations.

Another related concern has to do with the potential for information overload, and the ability of the subjects to process and interpret both the magnitude and complexity of information that could be provided pursuant to Kreuter's model. As a case in point, consider the challenge of conveying risk information to subjects regarding a single risk factor. One recommendation found in the literature is to convey such information to subjects in quantitative terms, preferably in the form of absolute risk by use of age intervals appropriate to the individual *(30–32)*. However, even this relatively straightforward recommendation regarding a single risk factor can present risk communicators with a significant challenge. Previous research has shown, for example, that when risk information is conveyed quantitatively, individuals tend to construct their interpretations in everyday life by use of qualitative categories, choosing to describe their risk, for instance, as being either high or low *(3)*. Reconciling this tendency of individuals to simplify their interpretations of risk with the sheer volume and complexity of information that might be conveyed by use of Kreuter's model is not a trivial issue, which will only be compounded if other factors (in addition to competing risks) are also introduced to subjects.

Despite the two concerns noted above, Kreuter has elevated our ongoing discussions regarding future directions for risk communication research. Thus, while Kreuter proposes several possible criteria for establishing priorities for risk reduction (e.g., epidemiologic risk, readiness to change, self-efficacy, objective difficulty in making specific behavior changes), exactly how these criteria might be combined most effectively for different populations is a legitimate and compelling question for future research. A related question is the extent to which priorities for risk reduction should be explicitly recommended versus allowing individuals to make these decisions themselves. There may even be a middle ground in which the goal of the intervention is to negotiate decisions with subjects regarding their priorities for behavior change. Taken together, these various options for involving subjects in the decision-making process, combined with the relative emphasis placed on the different prioritizing factors proposed by Kreuter, present an intriguing and provocative agenda for future research. However, while this agenda is both compelling and visionary, it should nonetheless be tempered by the state of the science and potential concerns regarding exportability, including the ability of individuals to process and interpret this information effectively. Given these concerns, one might conclude that Kreuter's research agenda exceeds what is feasible at the present time, especially within the context of large-scale phase III randomized trials in cancer prevention and control. If this is indeed true, it can hardly be viewed as a criticism, since all visions of the future should push the envelope of possibility beyond the reach of where we are today.

Selected Recommendations for Future Research

Ultimately, informed discussions regarding future directions for risk communication research must position specific recommendations within the larger context of other competing research questions. This must occur, given that there are finite resources available to support risk communication research in cancer prevention and control. Thus, by way of conclusion, listed below in no particular order are several suggestions in this regard.

More Research on the Presentation of Risk Information to Subjects

The field of risk communication would clearly benefit from more research on how best to convey risk information to specific target audiences, using both visual and text formats. The organizers of this conference are keenly aware of this issue, as evidenced by the paper by Lipkus *(33)*. In another conference paper, Rothman and Kiviniemi *(34)* have argued that risk information should be presented within the larger informational context of the antecedents and consequences of the targeted health problem. An additional issue that has caught my attention has to do with the recommendation to present quantitative descriptions of risk *(3,31,32,35)*, perhaps to the exclusion of qualitative presentations of risk *(30,36)*. However, as noted above, previous research has also shown that when risk information is conveyed in quantitative terms, individuals tend to construct their interpretations using ordinal or qualitative categories of risk *(3)*. If individuals are indeed prone to impose qualitative interpretations on quantitative risk information, perhaps we should acknowledge this fact and help individuals navigate through this process as part of our risk communication interventions. Thus, others have suggested that both quantitative and qualitative descriptions of risk should be presented to subjects *(37)*, using empirically based guidelines for translating quantitative risk into qualitative categories *(38,39)*. It would seem that more research is needed to identify optimal strategies for integrating both quantitative and qualitative presentations of risk into our risk communication interventions.

Establishing Guidelines for Measuring Risk

Even a superficial review of the literature reveals a fundamental lack of standardization with respect to measuring risk. Common formats include a wide variety of fixed-response categorical questions that may or may not provide a specific referent for comparative judgments by subjects, and various numerical scales that utilize a wide range of metrics for positioning subjects on a continuum of perceived risk. This lack of standardization clearly inhibits our ability to make comparisons of risk across studies and populations, and across time. It would

seem worthwhile for the NCI to convene a consensus workshop to make recommendations for standardization in this field of study, in a manner similar to the consensus workshops that have been convened for dietary assessments *(40)*. This reeommendation would seem to be especially timely, given that measures of perceived risk are being used extensively as intermediate outcomes, for end point assessments, and as key variables in our tailoring algorithms. However, it should be emphasized that what is not being suggested is that all future studies move forward in lock-step fashion using only those items that reflect standardization. Nonetheless, there would be benefit to establishing guidelines for a core set of risk assessment questions and formats that would be recommended across studies.

More Research Testing Interventions to De-Bias Optimistic and Pessimistic Perceptions of Risk

Various theories of behavior change highlight perceived risk as a major causal factor, including the Health Belief Model *(41)*, Precaution Adoption Model *(42,43)*, and Protection Motivation Theory *(44)*. Thus, one major challenge in risk communication arises from the observation that many people tend to underestimate their personal risk *(12,42,45–50)*, which, according to theory, should lower the probability of adopting precautionary or risk-reducing behaviors. This key premise provides the underlying rationale for targeting these optimistic perceptions of risk in our risk communication interventions.

Investigators have achieved varying levels of success with respect to de-biasing optimistic perceptions of risk *(12,29,34, 51)*. One promising line of research invokes social comparison theory to de-bias optimistic perceptions of risk. Thus, instead of providing risk information (for social comparison purposes) that is either general in nature or linked to "typical others," one might use "similar, individuated, physically present others as comparison targets" *(51)*. As noted by Klein and Weinstein *(51)*, such comparison targets could include similar peers who have a healthier lifestyle than the subject (and hence a lower risk), or they may include comparison targets at high risk and otherwise similar to the subject.

Although social comparison theory offers significant promise in de-biasing optimistic perceptions of risk, more research is needed. In particular, socioeconomic and cultural influences on the social comparison imperative need to be explored, as well as other maintenance strategies used by subjects to protect and sustain their risk perceptions. Other promising lines of research, some of which may directly or indirectly benefit attempts to de-bias optimistic perceptions of risk, include message framing *(52,53)*, targeting the attributions of risk *(42,54)*, and presenting risk information within a larger informational context that highlights both the antecedents and consequences of a potential health problem *(34)*. In addition, coping styles with respect to information preferences (e.g., information monitors versus blunters) might also help tailor the level and type of information provided to subjects *(55–58)*. These promising lines of inquiry, while not exhaustive of all possible opportunities for future research, nonetheless illustrate the dynamic nature of the state-of-the-science, and the potential to improve our ability to convey risk information to subjects.

Although receiving less attention in the research literature, efforts to de-bias pessimistic perceptions of risk should also be included in our agenda for future research. It is instructive to note in this regard that for some cancers (most notably breast cancer), a sizable percentage of individuals may overestimate their perceptions of risk *(59–63)*. If these pessimistic perceptions of risk result in a greater likelihood of adopting precautionary behaviors, then efforts to de-bias these pessimistic perceptions of risk may become less compelling and might even be viewed as counter-productive. This implicit dilemma is, perhaps, best illustrated with the following two questions. First, to what extent should we tolerate and accept pessimistic perceptions of risk, especially if they increase the probability of adopting precautionary behaviors in cancer prevention and control? And second, will pessimistic perceptions of risk, if taken to the extreme, impede behavior change by triggering denial and avoidance behavior? As will be noted below, the answers to these two questions should affect how we design and implement future risk communication programs.

More Research Examining Risk Communication as a Strategy for Behavior Change

Although various theories of behavior change highlight perceived risk as a major causal factor, what is not clearly understood is the relative efficacy of this theoretical construct in promoting behavior change, especially in comparison to other theoretical constructs in our models of health behavior. Under what conditions will manipulations of perceived risk, or more precisely, efforts to de-bias optimistic perceptions of risk yield the greatest impact in promoting behavior change? How does the relative efficacy of this theoretical construct differ by population subgroup or across different risk-reducing behaviors? The answers to these and similar questions have yet to be provided, which underscores the enormous challenges that lie ahead as we plot the future course of risk communication research in cancer prevention and control.

Of special note in this regard is the tendency for researchers to test interventions that simultaneously manipulate several theoretical constructs, thereby complicating efforts to disentangle and identify the specific theoretical mechanisms responsible for behavior change. The NCI is acutely aware of this problem, and thus has advocated that researchers devote more attention to specifying these mechanisms in their research *(64)*. This call to action will require appropriate modifications to our research designs, including differential manipulations of various theoretical constructs within our phase III trials. Although such studies may place a premium on the size of our samples, they are nonetheless needed if we are to advance the science of cancer prevention and control, and, more specifically, enhance our understanding of risk communication as a strategy for promoting cancer-protective behaviors nationwide.

As mentioned above, a related concern involves specifying the conditions under which risk information might impede behavior change. As noted at this conference, the available evidence would seem to support the hypothesis of a positive linear function between perceived risk and risk-reducing behaviors, at least for cancer screening *(65)*. Implicit in this empirical relationship is the notion that pessimistic perceptions of risk may facilitate the adoption of precautionary behaviors. Nonetheless, it is important to remember that this "best-fitting model" may not apply to specific individuals. Another model found in the literature suggests a curvilinear relationship between perceived risk and behavior *(66)*. According to this perspective, extreme perceptions of risk may exacerbate feelings of distress, and perhaps trigger denial and avoidance behavior, thereby producing a

curvilinear relationship between perceived risk and risk-reducing behaviors. This latter model takes on added significance, given that there is evidence suggesting that extreme perceptions of risk and/or elevated levels of cancer-specific distress may impede the adoption of precautionary behaviors in cancer prevention and control *(67–69)*.

For example, in a recent review of the literature, McCaul et al. *(70)* found that perceptions of breast cancer risk were predictive of breast cancer screening behavior in a positive linear fashion. However, virtually all of the studies included in this review involved samples drawn from the general population. It is conceivable that the hypothesized curvilinear relationship (described above) remains undetected in studies of the general population, given that extreme perceptions of risk may be relatively rare in such populations (thus truncating the distribution of perceived risk). In contrast, in samples of higher-than-average-risk populations, where concerns about a specific risk may be especially acute, risk perceptions may predict precautionary behaviors in a curvilinear fashion, especially if the behavior in question triggers anxiety or fear because of its potential to confirm these perceptions of risk, as in the case of breast cancer screening among populations at high risk for breast cancer.

Given the above, it would seem prudent in our research to identify at baseline those subjects who embrace extreme pessimistic biases of risk and elevated levels of cancer-specific distress. This cautionary note is being introduced for three main reasons. First, our risk messages should be appropriately tailored to this situation (e.g., instead of focusing on reducing optimistic perceptions of risk, our efforts should be directed in precisely the opposite direction). Second, and perhaps more importantly, we would not want to promote further denial or avoidance behavior or infliet additional distress by exacerbating extreme perceptions of risk. And finally, as suggested recently by Lerman et al. *(71)*, if cancer-specific distress is substantially elevated, efforts to counsel subjects about their risk may first need to address these psychosocial sequela.

Our colleagues in the field of genetic testing are keenly sensitive to the issue of psychosocial distress and thus have established various guidelines and procedures for minimizing serious psychosocial sequela attending disclosure of positive test results *(3,72,73)*. I would argue that we should be no less vigilant in other areas of risk communication involving groups at higher-than-average risk. Thus, we might consider assessing cancer-specific distress in all populations at higher-than-average risk, especially if elevated levels of distress are suspected, and we are attempting, as part of our intervention, to manipulate perceptions of risk. By routinely making these assessments at baseline and follow-up, we could also begin to assemble evidence regarding the prevalence of this potential problem in our research of high risk populations.

Finally, it seems apparent that the interrelationships between risk perceptions (especially extreme perceptions of risk), cancer-specific distress, and the adoption of precautionary behaviors need to be further clarified. For example, future research might conclude that extreme perceptions of risk is not what impedes behavior change or triggers denial and avoidance behavior, but rather its corollary, cancer-specific distress. This hypothesis suggests the need to model these factors in a multivariate environment, which again underscores the value of measuring both risk perceptions and cancer-specific distress in our studies of risk communication.

More Research Investigating the Sociology of Risk Communication

It would seem that risk communication research in cancer prevention and eontrol is currently preoccupied with a focus on the individual. This preoccupation may be a logical outgrowth of many of our theories of behavioral change, which tend to emphasize, as first-order causal factors, the knowledge base of individuals, their attitude and belief structures, and their perceptions of self-competency and various other facilitators and barriers to behavior change. When the role of one's larger social environment is acknowledged in our theories of health behavior, it is usually within the context of operating through these more proximal causal pathways. However, regardless of where these sociological influences are positioned in the causal chain, they need to be studied. More specifically, what is not fully understood is the role of sociocultural norms, and by extension, various significant others (and especially the family) in modifying or mediating risk communication on individual health behavior. However, such influences from one's larger social environment are undoubtedly present, perhaps quite compelling, and clearly in need of more research. Similarly, we lack a basic understanding of how individual risk information might impact the family, including familial relationships and the risk perceptions of other family members.

More Research That Specifically Targets Minority and Underserved Populations

In virtually all areas of inquiry in behavioral science, the first generation of research tends to focus on populations that are the most accessible, which typically translates into studies that focus on the majority (White/Anglo) population. However, as suggested by Huerta and Macario, the time has arrived for risk communication research to embrace underserved and minority populations. Given the NCI's commitment to promote research on underserved populations, this emphasis should be of high priority, perhaps facilitated with program announcements or RFAs that would specifically encourage such research.

More Research Testing Tailored Message Interventions, Including Risk Communication Interventions Within Diverse Community Settings

One of the most significant developments in recent years has been the introduction of computerized tailoring algorithms and interventions to promote behavior change in cancer prevention and control *(4–6,74)*. The capability of this technology to enhance the "self-relevance" of educational messages and print material could conceivably revolutionize the field of health education. As noted previously, more research is needed that tests this technology for risk communication within existing community settings. The NCI has been at the forefront in supporting such research, which needs to receive continued funding. Such research might be further encouraged through program announcements and RFAs that would specifically encourage partnerships with existing community intermediaries beyond the traditional health care system, especially as a strategy for reaching underserved populations.

Summary

The paper by Huerta and Macario reminds us of the need to consider sociocultural factors when designing risk communica-

tion interventions for minority and underserved populations. Their paper also reminds us that the time has arrived to focus more of our research efforts in risk communication on these special populations. In contrast, Dr. Kreuter challenges us to consider multiple and competing risks in cancer prevention and control, and offers a provocative framework for embracing this challenge. Clearly, both of these papers, combined with other priorities for research identified herein, illustrate the rich and compelling opportunities that exist as we continue to refine and improve the science of cancer prevention and control.

References

(1) Huerta EE, Macario E. Communicating health risk to ethnic groups: reaching Hispanics as a case study. Monogr Natl Cancer Inst 1999;25:23–26.

(2) Kreuter MW. Dealing with competing and conflicting risks in cancer communication. Monogr Natl Cancer Inst 1999;25:27–35.

(3) Bottorff JL, Ratner PA, Johnson JL, Lovato CY, Joab SA. Communicating cancer risk information: the challenge of uncertainty. Patient Educ Couns 1998;33:67–81.

(4) Skinner CS, Siegfried J, Kegler M, Strecher VJ. The potential of computers in patient education. Patient Educ Couns 1993;22:27–34.

(5) Rimer BK, Glassman B. Tailored communication for primary care settings. Methods Inf Med 1998;37:171–8.

(6) Strecher VJ, Greenwood T, Wang C, Dumont D. Interactive multimedia and risk communication. Monogr Natl Cancer Inst 1999;25:134–9.

(7) De Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: Introduction to a new approach. Patient Educ Couns 1999;36:99–105.

(8) Skinner CS, Strecher VJ, Hospers H. Physician recommendation for mammography: do tailored messages make a difference? Am J Public Health 1994;84:43–9.

(9) Strecher VJ, Kreuter MW, DenBoer DJ, Kobrin SC, Hospers HJ, Skinner CS. The effects of computer-tailored smoking cessation messages in family practice settings. J Fam Pract 1994;39:262–70.

(10) Campbell MK, DeVellis BM, Strecher VJ, Ammerman AS, DeVellis RF, Sandler RS. The impact of message tailoring on dietary behavior change for disease prevention in primary care settings. Am J Public Health 1994;84: 783–7.

(11) Campbell E, Peterkin D, Abbott R, Rogers J. Encouraging underscreened women to have cervical cancer screening: The effectiveness of a computer strategy. Prev Med 1997;26:801–7.

(12) Kreuter MW, Strecher VJ. Changing inaccurate perceptions of health risk: results from a randomized trial. Health Psychol 1995;14:56–63.

(13) Rakowski W, Ehrich B, Goldstein MG, Rimer BK, Pearlman DN, Clark MA, et al. Increasing mammography among women aged 40–74 by use of a stage-matched, tailored intervention. Prev Med 1998;27:748–56.

(14) Bental DS, Cawsey A, Jones R. Patient information systems that tailor to the individual. Patient Educ Couns 1999;36:171–80.

(15) Lowensteyn I, Joseph L, Levinton C, Abrahamowicz M, Steinert Y, Grover S. Can computerized risk profiles help patients improve their coronary risk? The results of the Coronary Health Assessment Study (CHAS). Prev Med 1998;27:730–7.

(16) Lipkus IM, Lyna PR, Rimer BK. Using tailored interventions to enhance smoking cessation among African-Americans at a community health center. Nicotine and Tobacco Research 1999;1:77–85.

(17) Rimer BK, Conaway M, Lyna P, Glassman B, Yarnall KSH, Lipkus I, et al. The impact of tailored interventions on a community health center population. Patient Educ Couns 1999;37:125–40.

(18) Eng E, Young R. Lay health advisors as community change agents. J Fam Community Health 1992;15:4–40.

(19) Eng E, Parker E, Harlan C. Lay health advisor intervention strategies: a continum from natural helping to paraprofessional helping. Health Educ Behav 1997;24:413–7.

(20) Campbell MK, Bernhardt JM, Waldmiller M, Jackson B, et al. Varying the message source in computer-tailored nutrition education. Patient Educ Couns 1999;36:157–69.

(21) Campbell MK, Demark-Wahnefried W, Symons M, Kalsbeek WD, Dodds J, Cowan A, et al. Improving fruit and vegetable consumption and cancer prevention: the Black Churches United for Better Health Project. Am J Public Health. In press.

(22) Tessaro I, Campbell M, Benedict S, Kelsey K, Heisler-MacKinnon J, Belton L, et al. Developing a worksite health promotion intervention: Health Works for Women. Am J Health Behav 1998;22:434–42.

(23) Marcus AC. The Cancer Information Service Research Consortium: a brief retrospective and a preview of the future. Prev Med 1998;27:S93–100.

(24) Burnett KF, Magel PE, Harrington S, Taylor CB. Computer-assisted behavioral health counseling for high school students. J Consult Psychol 1989;36:63–7.

(25) Brug J, Steenhuis I, van Assema P, DeVries H. The impact of a computer-tailored nutrition intervention. Prev Med 1996;25:236–42.

(26) Brug J, Glanz K, van Assema P, Kok G, Van Breukelen GJP. The impact of computer-tailored feedback and interactive feedback on fat, fruit and vegetable intake. Health Edu Behav 1998;25:357–71.

(27) Brug J, Steenhuis IH, van Assema P, Glanz K, De Vries H. Computer-tailored nutrition education: differences between two interventions. Health Educ Res 1999;14:249–56.

(28) Campbell MK, Honess-Morreale L, Farrell D, Carbone E, Brasure M. A tailored multimedia nutrition education pilot program for low income women receiving food assistance. Health Educ Res 1999;14:257–67.

(29) Weinstein ND, Klein WM. Resistance of personal risk perceptions to debiasing interventions. Health Psychol 1995;14:132–40.

(30) Stefanek ME. Counseling women at high risk for breast cancer. Oncology 1990;4:27–33.

(31) Vogel VG, Yeomans A, Higginbotham E. Clinical management of women at increased risk for breast cancer. Breast CA Res Treat 1993;28:195–210.

(32) Bryant HE, Brasher PM. Risks and probabilities of breast cancer: short-term versus lifetime probabilities. Can Med Assoc J 1994;150:211–16.

(33) Lipkus IM, Hollands JG. Visual communication of risk. Monogr Natl Cancer Inst 1999;25:149–63.

(34) Rothman AJ, Kiviniemi MT. Treating people with information: an analysis and review of approaches to communicating health risk information. Monogr Natl Cancer Inst 1999;25:44–51.

(35) Love RR. The article reviewed (comments on paper by ME Stefanek: Counseling women at high risk for breast cancer). Oncology 1990;4:37–8.

(36) Kelly P. Risk counseling for relatives of cancer patients: New information, new approaches. J Psychosocial Oncology 1987;5:65–79.

(37) Holtgrave DR, Tinsley BJ, Kay LS. Encouraging risk reduction: a decision-making approach to message design. In: Mailbach F, Parrott RL, editors. Designing health messages. Thousand Oaks (CA): Sage Publications; 1995. p. 24–40.

(38) Mosteller F, Youtz C. Quantifying probabilistic expressions. Statistical Science 1990;5:2–12.

(39) Kadane JB. Comment: quantifying probabilistic expressions. Statistical Science 1990;5:18–20.

(40) National Center for Health Statistics (NCHS), Division of Health Examination Statistics. Consensus workshop on dietary assessment: nutritional monitoring and tracking the Year 2000 objectives. Wright JD, Ervin B, Briefel RR, editors. February 1993.

(41) Becker MA, editor. The health belief model and personal health behavior. Health Educ Monogr 1974;2.

(42) Weinstein ND. Why it won't happen to me: perceptions of risk factors and susceptibility. Health Psychol 1984;3:431–57.

(43) Weinstein ND. The precaution adoption process. Health Psychol 1988;7: 355–86.

(44) Rogers RW. Cognitive processes in fear appeals and attitude change: a revised theory of protection motivation. In: JT Cacioppo, RE Petty, editors. Social Psychophysiology. New York (NY): Guilford Press; 1983. p. 153–76.

(45) Avis NE, Smith KW, McKinlay JB. Accuracy of perceptions of heart attack risk: what influences perceptions and can they be changed? Am J Public Health 1989;79:1608–12.

(46) Becker DM, LeVine DM. Risk perception, knowledge, and lifestyles in siblings of people with premature coronary disease. Am J Prev Med 1987; 1:45–50.

(47) Strecher VJ, Kreuter MW, Kobrin SC. Do cigarette smokers have unrealistic perceptions of their heart attack, cancer and stroke risks? J Behav Med 1995;18:45–54.

(48) Weinstein ND. Unrealistic optimism about future life events. J Pers Soc Psychol 1980;39:806–20.
(49) Weinstein ND. Unrealistic optimism about susceptibility to health problems. J Behav Med 1982;5:441–60.
(50) Weinstein ND. Reducing unrealistic optimism about illness susceptibility. Health Psychol 1983;2:11–20.
(51) Klein WM, Weinstein ND. Social comparison and unrealistic optimism about personal risk. In: Buunk BP, Gibbons FX, editors. Self, coping and well-being: perspectives from social comparison theory. Mahwah (NJ): Lawrence Erlbaum; 1997. p. 25–61.
(52) Rothman AJ, Salovey P, Antone C, Keough K, Martin CD. The influence of message framing on intentions to perform health behaviors. J Experim Soc Psychol 1993;29:408–33.
(53) Banks SM, Salovey P, Greener S, Rothman AJ, Moyer A, Beauvais J, et al. The effects of message framing on mammography utilization. Health Psychol 1995;14:178–4.
(54) Blalock SJ, Devellis BM, Afifi RA, Sandler RS. Risk perceptions and participation in colorectal screening. Health Psychol 1990;9:792–806.
(55) Miller SM. Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. J Pers Soc Psychol 1987;52: 345–53.
(56) Miller SM, Combs C, Stoddard E. Information, coping and control in patients undergoing surgery and stressful medical procedures. In: Steptoe A, Appels A, editors. Stress, personal control, and health. New York (NY): John Wiley & Sons; 1989. p. 107–19.
(57) Schwartz MD, Lerman C, Miller SM, Daly M, Masny A. Coping disposition, perceived risk, and psychological distress among women at increased risk for ovarian cancer. Health Psychol 1995;14:232–5.
(58) Lerman C, Schwartz MD, Miller SM, Daly M, Sands C, Rimer BK. A randomized trial of breast cancer risk counseling: interacting effects of counseling, educational level and coping style. Health Psychol 1996;15: 75–83.
(59) Black WC, Nease RF, Tosteson AW. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst 1995;87:720–31.
(60) Bondy ML, Vogel VG, Halabi S, Lustbader ED. Identification of women at increased risk for breast cancer in a population-based screening program. Cancer Epidemiol Biomarkers Prev 1992;1:143–7.
(61) Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med 1991;114:657–61.
(62) Evans DG, Burnell LD, Hapwood P, Howell A. Perceptions of risk in women with a family history of breast cancer. Br J Cancer 1993;67:612–4.
(63) Skinner CS, Kreuter MW, Kobrin S, Strecher VJ. Perceived and actual breast cancer risk. J Health Psychology 1998;3:181–93.
(64) Lerman C, Rimer B, Glynn T. Priorities in behavioral research in cancer prevention and control. Prev Med 1997;26(Suppl):S3–9.
(65) McCaul KD, Tulloch HE. Cancer screening decisions. Risk Communication. Monogr Natl Cancer Inst 1999;25:52–8.
(66) Hailey BJ. Family history of breast cancer and screening behavior: an inverted U-shaped curve? Med Hypotheses 1991;36:397–403.
(67) Lerman C, Daly M, Sands C, Balshem A, Lustbader E, Heggan T, et al. Mammography adherence and psychological distress among women at risk for breast cancer. J Natl Cancer Inst 1993;85:1074–80.
(68) Lerman C, Schwartz M. Adherence and psychological adjustment among women at high risk for breast cancer. Breast Cancer Res Treat 1993;28: 145–55.
(69) Kash KM, Holland JC, Harper MS, Miller DG. Psychological distress and surveillance behaviors of women with a family history of breast cancer. J Natl Cancer Inst 1992;84:24–30.
(70) McCaul KD, Branstetter AD, Schroeder DM, Glasgow RE. What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. Health Psychol 1996;15:423–9.
(71) Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995;87:286–92.
(72) MacDonald KG, Doan B, Kelner M, et al. Genetic vulnerability: the unwanted inheritance. In: Taylor KM, DePetrillo D, editors. Critical choices: ethical, legal and sociobehavioral implications of heritable breast, ovarian and colon cancer. Toronto (Canada): International Research and Policy Symposium; 1995. p. 19–60.
(73) Mulvihill JJ. Genetic counseling of the cancer patient. In: Devita VT Jr, Hellman S, Rosenberg SA, editors. Cancer: principles and practices of Oncology. Vol. 2. 4th ed. Philadelphia (PA): Lippincott; 1992. p. 2529–37.
(74) Voelker R. Personalized messages invite more mammograph. JAMA 1994; 86:17.

Introduction of Section: Enabling Informed Decisions About Cancer Risk

Barbara Bowles Biesecker

Determining health risk is a science. Communicating risk information to a patient remains an art. This health care practice depends on complexities of human relationships, communication styles, intellectual abilities, and emotional receptivity. For the practice of risk communication to meet its goals, this art must also be informed by science. Evidence of how risk communication is most successful and when it translates into increased health should become the basis for the future practice of risk communication.

Research on individual risk perception and health care decisions concerning cancer risks can currently inform the practice of risk communication. The focus of the next section in the monograph addresses risk perception and decision making under conditions of uncertainty, particularly where the information is complex, incomplete, and changing. While valuable data are reviewed and synthesized, many important areas for future research remain.

My own profession of genetic counseling provides perhaps no better example of the need for health care providers to understand individual risk perception and the complexities of making health-related decisions. Genetic counselors appreciate the subjective nature of genetic risk. The loss inherent in a genetic risk is peculiar to the person concerned, as are the significance of the loss and the chance of its occurrence. Counselors often note the ways their clients polarize risk—its significance reduced to the dichotomization that either the genetic condition will or will not occur. Overall, it remains uncertain in genetic counseling how clients interpret and use genetic risk information in making their decisions about genetic testing and the value they place on the information. Studies of cancer risk communication have now been conducted in the context of genetic testing and will begin to shape future genetic counseling practices. The theoretical frameworks for understanding and communicating risk information and the efficacy of various decision-making models need to be studied not only in genetic counseling but also to a greater degree in cancer risk communication.

The research on individual cancer risk communication related to risk perception and decision making reviewed in this monograph should inform future research and policy. The following background papers raise interesting areas of future inquiry. The first review article, by Alexander Rothman and Marc Kiviniemi, synthesizes the research evaluating approaches to communicating health risk information. They suggest that risk communication that includes not only a numerical risk estimate but also information about the causes and consequences of a health problem may prove most effective in successfully communicating risk information. Kevin McCaul and Heather Tulloch evaluate the evidence for why people decide to obtain or avoid cancer screening. They conclude that counseling should focus not only on removing barriers to screening but also on influencing beliefs about the value of screening. Robert Croyle and Caryn Lerman summarize results from recent studies of offering genetic testing for cancer predisposition to high-risk families. The evidence of efficacy and the policy implications of research findings on the use of decision aids for patients considering health care options are reviewed by Annette O'Connor and colleagues. They recommend that practitioners identify which health care decisions are sensitive to patients' values and become familiar with available resources. Howard Leventhal and Barnett Kramer have authored commentaries on the articles.

Affiliation of author: Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD.

Correspondence to: Barbara B. Biesecker, M.S., National Institutes of Health, Bldg. 10, Rm. 10C101, Bethesda, MD 20892-1852 (e-mail: barbarab@nhgri.nih.gov).

Treating People With Information: an Analysis and Review of Approaches to Communicating Health Risk Information

Alexander J. Rothman, Marc T. Kiviniemi

The communication of risk information is a fundamental aspect of nearly all health promotion interventions. However, no consensus exists regarding the most effective way to provide people with risk information. We will review and evaluate the relative merits of two approaches to risk communication. One approach relies on the presentation of numerical information regarding the probability of a health problem occurring, whereas the other relies on the presentation of information about the antecedents and consequences of a health problem. Because people have considerable difficulty understanding and using quantitative information, the effectiveness of interventions that rely solely on numerical probability information has been limited. Interventions that provide people with a broader informational context in which to think about a health problem have had greater success systematically influencing perceptions of personal risk but have several important limitations. However, before any final conclusions can be drawn regarding the relative merits of different communication strategies, investigators must agree on the specific criteria that should be used to identify an effective intervention. [Monogr Natl Cancer Inst 1999;25:44–51]

Because of the influence that perceptions of personal risk are thought to have on people's health practices *(1,2)*, the communication of health risk information has become a prominent part of health prevention efforts. People are frequently provided with numerical information about the probability that a specific pattern of behavior (e.g., smoking) will lead to a particular health problem (e.g., lung cancer). Although the dissemination of this information has increased public awareness of potential health problems, its effect on perceptions of personal risk for these problems has been less dramatic *(3)*.

Why has this information had such a limited effect on people's beliefs about their own personal vulnerability? A possible explanation is that people routinely misunderstand numerical information about the probability that an outcome will occur *(4)*. This confusion leads people to misinterpret the implications of the information and, thus, undermines any systematic effect it might otherwise have on their beliefs and behavior. Moreover, people are not passive, unbiased processors of information about their health status *(5)*. They welcome favorable information about their health but often engage in strategies that minimize or discount unfavorable health information—strategies that, in turn, blunt any influence this information might have on decision making and behavior [e.g., *(6–8)*].

In light of these cognitive and motivational barriers, what is the best way to communicate health risk information so that people both understand it and recognize its implications for their own risk status? We will examine several methods that have been used to communicate health risk information and evaluate their effect on perceptions of personal risk. Our goal is to identify conditions that maximize the desired impact of health risk information on people's health beliefs. The degree to which risk-based interventions affect people's behavioral practices is reviewed elsewhere in this volume (*see* Gerrard et al. [this monograph] and McCaul and Tulloch [this monograph]).

Before reviewing the merits of different approaches to risk communication, we consider two issues that are fundamental to any assessment of risk communication strategies. First, how do people think about and form judgments of health risks? Interventions designed to help people assess their risk will be successful only if consideration is given to how people process different types of health risk information. Second, what is the most appropriate way to assess the impact of a health risk intervention? An intervention could influence knowledge about a risk factor, beliefs about personal risk, perceptions of how one's own risk compares with the average person's risk, or intentions to act to reduce one's risk. Because these different measures can provide quite different impressions of people's beliefs, the specific set of criteria an investigator chooses to rely on can affect whether or not a particular intervention is perceived to be effective *(3,9)*.

How Do People Think About Health Risks?

People are quite willing to provide estimates of personal health risks, but what these estimates reveal about their health beliefs is unclear. Although risk estimates are typically interpreted as the probability with which people believe a health problem will occur, there is considerable evidence that they reflect a broader set of cognitive and affective beliefs. Analyses of how people interpret and compare health risks have consistently revealed that lay perceptions of risk are not simply a function of probability information but rather are affected by the beliefs people hold about both the antecedents of the problem (for example: "Is it voluntary?" "Is it controllable?") and its consequences [for example: "Can it be detected?" "Would it be catastrophic?"; *see (10,11)*].

Why might people's perceptions of risk be affected by information about the antecedents and consequences of a health problem? First of all, people think about health problems in relation to their causes and consequences and, in fact, health information is organized in memory primarily around these two classes of

Affiliation of authors: Department of Psychology, University of Minnesota, Minneapolis.

Correspondence to: Alexander J. Rothman, Ph.D., Department of Psychology, University of Minnesota, Elliott Hall, 75 East River Rd., Minneapolis, MN 55455 (e-mail: rothm001@umn.edu).

See "Notes" following "References."

© Oxford University Press

information *(12)*. Moreover, some investigators have proposed that people come to understand their risk on the basis of their ability to mentally simulate or imagine themselves experiencing the problem *(13,14)*. Thus, information will affect judgments of risk to the extent that it enables one to imagine that "it could happen to me." In the context of this framework, knowledge of both the antecedents and the consequences of a health problem has clear informational value.

Antecedent information helps people think about *how* a health problem could develop. It renders salient the specific factors that either promote or prevent the development of a health problem [e.g., *(15)*]. Moreover, it helps people recognize the links between the very things that they do—or do not do—and unwanted health outcomes [e.g., *(16)*]. Whereas information about the antecedents of a health problem underscores how people can alter their chances of developing a health problem, information about consequences emphasizes what it would be like to actually have the health problem. Vivid information about the consequences of the health problem helps people to recognize the severity of *what* could happen to them [e.g., *(17,18)*] and, moreover, that it actually has happened to people just like them [e.g., *(19)*]. Thus, both types of information provide useful and nonoverlapping material to help individuals assess and interpret their risk.

Even though numerical probability information provides people with a precise description of their risk, its informational value is unclear. In and of itself, providing people with numerical probability information about a health problem (for example: "You have a one in 500 chance of developing breast cancer") is likely to have little effect on their ability to imagine either *what* might happen to them or *how* it might occur. In fact, research has shown that people tend to underutilize probability information except when it provides explicit causal information about how a particular event will occur *(20,21)*. Given that particular types of information about a health problem may have greater informational value than others, any evaluation of health risk communications must consider the characteristics of the specific information provided. In particular, messages that exclusively provide people with numerical probability information need to be differentiated from those that provide people with more general information about the antecedents or consequences of a health problem.

Evaluating the Impact of Health Risk Information

How does one determine whether a health risk message has been effective? Investigators have assessed people's knowledge of the specific risks posed by a behavior (for example: "How does smoking affect people's risk for lung cancer?"), beliefs about the absolute magnitude of their own risk (for example: "Given your current smoking habit, what is your risk for developing lung cancer?"), and beliefs about how their risk compares with that of a specified comparison group (for example: "How does your risk for developing lung cancer compare with that of the typical smoker?").

Evidence that people are aware of the risks posed by a particular pattern of behavior may indicate the successful communication of health risk information, but it cannot tell us whether people recognize the risk as personally relevant. Direct measures of personal risk are needed to answer this question. However, it is not clear whether the criterion for an effective intervention should be based on changing absolute levels of personal risk or in modifying comparative risk. Numerous studies have revealed that people report their own risk of experiencing a health problem to be less than that of the average person and that this occurs even when they are given information about the average person's risk or behavior [e.g., *(22,23)*]. This optimistic bias has been taken to indicate that people systematically underestimate important personal health risks and, thus, considered a barrier to the adoption of precautionary behaviors [*(24)*; but *see (13)*]. However, the absence of any empirical evidence that people's behavioral decisions are related differentially to perceptions of absolute and comparative risk limits any conclusion that can be drawn as to which measure offers the most informative assessment of people's beliefs.

Different Approaches to Communicating Health Risk Information

We now turn to a brief review of the literature on communicating health risk information. The review covers studies that have systematically compared different methods of communicating risk information, all of which were developed to affect people's beliefs about health risks that were associated with either their own behavior (e.g., smoking or sexual behavior) or their environment (e.g., radon). None of the studies included in this review addressed the communication of risks posed by different medical treatments.

We distinguish between two general approaches to communicating health risk information: a numerical probability-based approach and a contextualized approach. A probability-based approach focuses on the presentation of numerical information regarding the probability of a given risk occurring. The primary goal of studies that have utilized this approach has been to identify the most effective way to communicate risk information. Specifically, these studies seek to determine how different presentation formats affect people's ability to understand and, in some cases, use the numerical information provided. These interventions have ranged from the straightforward communication of a numerical probability (for example: "You have a one in 500 chance of developing cancer in the next 10 years") to presentation formats designed to heighten the salience or impact of the statistical information. Some of the more innovative presentation strategies have involved framing the information in terms of how frequently someone will die in a community as a result of a given risk *(25)* or presenting risk information in such a way that it can be compared with the magnitude of other risks *(26)*.

A second approach to risk communication has focused on providing people with an informational context in which to understand and interpret their risk. The defining feature of a contextualized approach is that it provides people with information about the antecedents and/or the consequences of a potential health problem. A variety of intervention methods can be considered contextual. Some interventions have emphasized the antecedents of a health problem by providing people with specific information regarding the link between a health problem and their own prior behavior or medical history [e.g., *(16)*]. Other interventions have used testimonials from people who have experienced the health problem to convey information about how the problem developed and/or how it has affected their lives [e.g., *(19,27,28)*]. Other contextualized interventions have fo-

cused on providing people with information or images that render the negative consequences of the health risk salient and vivid [e.g., *(29,30)*]. In almost every case, the primary goal of the intervention has been to increase people's recognition of the risk posed by their behavior or environment. Thus, these interventions have been developed not only to communicate risk information, but also to do it in as persuasive a manner as possible.

There are several criteria that an investigator can use to determine the effectiveness of any given communication strategy. First, an effective strategy might lead people to be better informed about the likelihood of a health problem. Interventions that emphasize the presentation of probability information have focused primarily on this criterion. Alternatively, an effective strategy could appropriately alter people's perceptions of personal or comparative risk. Third, an effective strategy could be one that heightens people's interest in relevant primary prevention behaviors, independent of changes in risk perceptions. Empirical tests of contextually based interventions have focused primarily on the latter two criteria. Whenever possible, we report the effect an intervention method has had on both absolute and comparative measures of risk as well as whether changes in risk perceptions were associated with changes in other risk-relevant beliefs.

Impact of Probability-Based Approaches

When investigators provide numerical probability information, they rely on the premise that people will respond to a given probability in a consistent manner. That is, a 10% risk should be interpreted as a 10% risk, regardless of whether it is presented as a percentage or an odds ratio or whether it is presented numerically or pictorially. Unfortunately, empirical tests of probability-based communications have typically failed to support this premise [e.g., *(31,32)*]. People have difficulty recalling probability information accurately [e.g., *(33,34)*], they make errors when asked to transform percentages into proportions or vice versa (e.g., Lipkus I: unpublished data [numeracy survey], 1998), and they confuse information about the frequency of a event with its rate of occurrence [e.g., *(35)*]. Thus, it would appear that people cannot reliably understand and interpret numerical probability statistics.

In fact, there is direct evidence that people do not appreciate some of the fundamental assumptions underlying probabilities. For example, Black et al. *(36)* reported that a substantial number of women estimated that their risk of developing breast cancer was less than their risk of developing *and* dying from it. People also have difficulty understanding and estimating cumulative risks associated with repeated patterns of behavior [e.g., contraception use *(37,38)*]. More broadly, Schwartz et al. *(39)* assessed people's familiarity with basic probability and numerical concepts (i.e., numeracy). They found not only that people's understanding of these mathematical concepts was quite poor, but also that their ability to accurately interpret numerical information regarding breast cancer and mammography was systematically dependent on their level of numeracy.

People's beliefs are also sensitive to the specific numbers used to illustrate a risk. For example, people may be more concerned about a potential health problem when they are told that a pattern of behavior will increase their risk of dying by 30% than if they are told that it will change their risk of dying from one in 10 000 to 1.3 in 10 000 *(11)*.[1] Several studies have shown that people respond differentially to a ratio when it is expressed as 1 : 10 versus 10 : 100 and that this occurs even when they understand that the two ratios illustrate the same probability *(40,41)*. Moreover, Yamagishi *(42)* found that people rated a health problem as riskier when they were informed that it kills 1286 of 10 000 people (12.86%) compared with 24.14 of 100 people (24.14%). It would appear that people fail to consider the relevant sample size when processing information regarding the number of people who have developed or died from a health problem [*but see (31)*].

In addition to the difficulties people have with the mathematical properties of probabilities, investigators must grapple with the fact that people find it difficult to ascribe meaning to a given level of risk. What does it mean to be told that you have a 1% chance of having a disabled child or that your risk of dying from cancer has increased from one in 10 000 to 1.3 in 10 000? Although these statistics may be extremely precise, people's responses to this information are anything but precise. When provided with numerical estimates, people appear to spontaneously transform the probability information into discrete categories [e.g., high or low risk *(43,44)*]. To the extent that people make decisions based on the belief that their risk is either high or low, the specific numerical probabilities they provide may not be particularly reliable or informative. Consistent with this analysis, Windschitl and Wells *(45)* found that people's numerical probability estimates failed to predict their preferences and behavioral intentions, whereas the verbal labels they used to describe the likelihood of an outcome consistently predicted the same set of decisions.

Because people have difficulty ascribing meaning to a probability estimate, they are acutely sensitive to available comparison information. In fact, when asked what information would help them understand a health risk, individuals typically request comparisons between the probability of different risks *(46)*. Evidence that people readily construe risks in a comparative frame comes from studies that have manipulated the comparison information provided. Klein *(47)* has shown that people's affective reactions to information about their risk depend on how their level of personal risk compares with the average risk, independent of their absolute level of personal risk. Studies that have presented risks along a continuum have shown that people's interpretation of the same numerical probability depended on whether it was placed at the high end or the low end of the continuum *(26)*. People's interpretation of the implications of their behavior is similarly affected by whether a response format leads them to believe that the frequency of their behavior is above or below the perceived norm (Rothman AJ, Haddock G, Schwarz N: manuscript submitted for publication). Of course, the manner in which risk information is presented is but one source of information that can affect how people interpret or use a given probability. Research has shown that people's knowledge, personality, and goals can also influence their interpretation of what a particular risk means [e.g., *(48–50)*].

The primary appeal of health risk communications that present numerical probability information is that they offer people precise information regarding the probability that a health problem will occur. However, the effectiveness of this communication approach is constrained by people's inability to accurately interpret and use numerical probabilities. In fact, the very strategies that people employ to render probability information more useful strip it of the very precision that this intervention approach offers.

Impact of Contextualized Approaches

In light of the limitations of a probability-based approach to communicating risk information, we now consider an alternative method that provides people with information to assist them in understanding the personal implications of a given health risk. Recall that, although risk is typically defined as the numerical probability of a given event occurring, lay conceptions of risk are based on a much richer set of cognitive and affective beliefs. When a health problem is being considered, information about the numerical probability of a risk is viewed as insufficient; people are interested in information about what causes a health problem, about the severity of its consequences, and about what can be done to either prevent or treat the problem *(51,52)*. This information is considered valuable because it enables people to develop a mental model that delineates the personal relevance of a given risk—what might cause one to develop a health problem and the potential costs of developing that problem.

Antecedents. Interventions that heighten the salience of factors that place people at risk for adverse outcomes have typically involved directing their attention to relevant aspects of their behavior or personal attributes. For instance, Gerrard et al. *(53)* had female Marines review their sexual and contraceptive histories prior to assessing their risks for pregnancy and human immunodeficiency virus (HIV) infection. This behavioral review heightened judgments of personal risk for HIV infection and, in some cases, personal risk for pregnancy. Other interventions have provided individuals with personalized information about the link between a health problem and specific behavioral or personal characteristics. This strategy has been somewhat successful in rendering both optimistic and pessimistic misperceptions about comparative risk more accurate *(16)*, although other investigators *(54)* have found that directing people's attention toward potential risk factors failed to reduce optimistic perceptions of comparative risk.

Some investigators have emphasized the need to differentiate between the absolute amount of risk-relevant information people can bring to mind and the relative ease with which that information is recalled. Traditionally, interventions have assumed that the more risk-increasing factors people bring to mind, the more at risk they will feel. However, the experienced ease with which information comes to mind has also been shown to systematically affect judgment [e.g., *(55)*]. Specifically, people who have easily brought a few risk-increasing factors to mind may infer a greater degree of personal risk than do people who had difficulty bringing to mind a larger number of risk factors. Both Rothman and Schwarz *(15)* and Raghubir and Menon *(56)* have successfully used this approach to heighten perceptions of personal risk, although only Raghubir and Menon obtained a reduction in participants' optimistic perceptions of comparative risk.[2] Perceptions of personal risk are also sensitive to the relative ease or difficulty with which risk-decreasing factors come to mind. In this case, having had difficulty generating a large number of risk-decreasing factors can lead to increased perceptions of personal risk *(15)*.

Consequences. Investigators have developed two different strategies to increase the salience of outcomes associated with a given health problem. In each case, the goal of the intervention has been to increase the personal relevance of the risk by helping people to recognize that the health problem could happen to them. Sherman et al. *(57)* provided people with a set of symptoms that were either easy or difficult to bring to mind and observed that the ease with which people could imagine themselves experiencing the symptoms of a disease heightened their perceptions of personal vulnerability. More recently, interventions have tried to emphasize the personal relevance of a health problem by highlighting the similarities between members of the targeted audience and people who have had personal experience with a health problem. For example, the presentation of personal testimonials by HIV-positive individuals has led to increased perceptions of personal risk—but only when people perceived themselves to be similar to the person providing the testimony [e.g., *(27,28); see also (58)*]. In fact, Gump and Kulik *(28)* have demonstrated that the level of personal risk people will recognize depends on the degree to which they are able to see themselves as different from the infected person.

A second approach to highlighting the outcomes associated with a given risk has involved messages that graphically illustrate the severity of the consequences associated with a health problem [*(8,18); see* Salovey et al. *(1)* for a recent review]. For example, Sutton and Hallett *(30)* successfully raised perceptions of personal risk related to seat belt use by showing people graphic scenes of car crashes and mortuaries. Similar effects have been found for images that depict outcomes related to cigarette smoking [e.g., *(29,59)*]. Although there is no empirical support for the premise that the disturbing nature of these appeals would lead people to minimize the personal relevance of a given health threat, people must know how they can respond to the health threat if a fear appeal is going to elicit a change in behavior *(18,60)*.

Given the findings obtained across a range of experimental paradigms, it would appear that interventions that render either the antecedents or the consequences of a health problem salient can have a meaningful effect on perceptions of personal risk. Although people may be responsive to information about what could happen to them and how it might occur, the specific processes by which this information affects risk beliefs have yet to be well specified. We consider this issue as well as potential limitations to this communication strategy in the following section.

What Have We Learned and What Do We Still Need to Know?

On the basis of the empirical research covered in our review, we believe that several conclusions can be drawn regarding the communication of health risk information. First, people do not respond in a consistent manner to communication strategies that rely solely on the presentation of numerical risk information. From the perspective of an expert, the precision afforded by a numerical probability estimate may be appealing, but its influence on lay people's beliefs and behavior is anything but precise. Any systematic impact that this information might have on decision making and behavior is undermined by people's inability to understand the meaning of a specific probability and their resulting computational errors. In an attempt to ascribe meaning to a given risk, people's inferences are strongly affected by information that is accessible at the time of judgment. Thus, there is considerable variability in how people respond to probability-based risk messages.[3] Interventions that provide people with information about the causes or consequences of a particular health problem appear to have a more consistent influence on perceptions of personal risk precisely because they systematically control the information people have available when draw-

ing inferences about their personal risk. Moreover, the observation that contextually based approaches to providing risk information are more effective is consistent with the broader finding that people are more likely to base their judgment on concrete, case-based information than on abstract, statistical information [e.g., *(63,64)*].

Recognizing the Limitations of a Contextually Based Communication Strategy

Although interventions that have utilized a contextually based approach to communicating risk information have had some success affecting perceptions of personal risk, investigators cannot assume that people will always respond appropriately to the health information provided. Because people want to maintain a favorable impression of their health status, they may, if possible, selectively attend to information about their health. For example, when asked to consider how their own risk for a health problem compares with that of the average person, people selectively bring to mind health practices that make them look good *(65)*. In a similar manner, people have a better memory for behavioral guidelines that place their behavior in a favorable (healthy) light than for those that place it in an unfavorable (unhealthy) light *(66)*. Finally, when asked to consider their own behavioral practices, people may selectively focus on the preventive actions that they have adopted rather than on the risks that they have taken [*(53,67)*; for a complete discussion of these findings, *see* Gerrard et al., this monograph].

Even when people do attend to information about potential health risks, they may search for flaws in the message or adopt higher standards for evaluating the quality of the information *(6,8,68)*. In situations where risk information cannot be refuted directly, people may adjust other beliefs in an attempt to counter the undesirable implications of the risk message *(7,69,70)*. For example, Gerrard et al. *(7)* observed that, even though young adults increased their perceptions of risk in response to an increase in their risk behaviors, they also increased their assessment of the prevalence of these behaviors (thereby normalizing the practice) and decided that the health implications of these behaviors were less relevant to future behavioral decisions.

However, the importance of people's attempts to minimize the implications of the risk information is unclear. Several studies have observed that, even when people act to minimize a health threat, they remain interested in information about how to address the problem [e.g., *(71–73)*]. For example, people who learned that they had borderline high cholesterol perceived high cholesterol to be a less serious health problem than did those who were told that they had low cholesterol, but they still expressed greater interest in taking steps to lower their cholesterol *(72)*. There is even evidence that interventions can prompt changes in behavioral intentions independent of observed changes in risk perceptions. Gump and Kulik *(28)* found that the HIV status of a man describing his experiences influenced intentions to be HIV tested independent of changes in perceptions of personal risk. Similarly, Evers et al. *(27)* found that heterosexuals reported stronger intentions to be HIV tested regardless of the sexual orientation of an HIV-positive speaker, whereas they changed their perceptions of personal risk only after having listened to a heterosexual speaker.

The observed dissociation between people's risk beliefs and their behavioral intentions should not be interpreted as a sign that changes in people's behavior can be elicited in the absence of information about health risks. Information about a potential health risk may be necessary to initiate the self-regulatory processes that underlie an appropriate behavioral response. However, people may independently regulate their affective and behavioral responses to a health threat, as suggested by Leventhal's Parallel Response Model *(12,18)*. Specifically, people may choose to respond both in ways that alleviate their distress about a potential health problem (thereby attenuating any change in perceived risk) and in ways that serve to reduce the likelihood that they might develop the health problem (thereby eliciting changes in behavior). It is precisely because people may respond to health risk information in myriad ways that investigators who test interventions must attend not only to changes in people's perceptions of the chance of a health risk occurring, but also to changes in people's beliefs about the implications of that risk and how they intend to respond to that risk. In the absence of an assessment of these ancillary beliefs, it may be difficult to draw conclusions regarding the practical significance of any observed change (or lack thereof) in perceived risk.

Moving Beyond the Perceived Magnitude of One's Risk

At present, health risk interventions are primarily evaluated on the basis of their ability to influence perceptions of risk. Although the possibility that behavioral intentions will change independent of any change in perceived risk indicates that investigators need to employ a broad array of indicators when evaluating an intervention, consideration should also be given to the specific manner in which changes in perceived risk are assessed. In fact, the apparent dissociation between risk beliefs and behavioral intentions may reflect, in part, the ways that investigators have operationalized perceived risk. Almost all measures of risk perception have been designed to detect changes in the perceived likelihood of a person's risk. However, what if an intervention is effective not because it leads people to believe something is more likely to happen but instead because it causes them to be more concerned about a given risk? For example, a person might still report a 1% risk of developing a health problem but be considerably more concerned about that level of risk. Current measures of risk perceptions are not able to detect changes in the meaning people ascribe to a specific level of risk. The inclusion of measures of worry and concern about a health risk might help address this problem. In fact, McCaul et al. *(74)* have found that how worried a woman was about developing breast cancer predicted mammography screening utilization even after controlling for her perception of personal risk. It might also be useful to assess whether an intervention has altered the salience or accessibility of a person's risk perceptions. Interventions might prove to be effective not because they increase people's perceptions of risk but because they increase the probability that beliefs about personal risk will come to mind when people are faced with a behavioral decision.

Toward a Model of Health Risk Communication

Despite the fact that investigators have assessed the merits of a range of risk communication strategies, our limited understanding of how different aspects of a health risk message affect people's beliefs and behaviors constrains any recommendations that can be formulated. The observation that information regarding both the antecedents and consequences of a health problem can reliably affect people's risk beliefs may offer a base on which a conceptual model of risk communication can be built.

Information about the antecedents and consequences of a health problem is believed to be effective because it enables people to imagine that "it could happen to them" *(13,14)*. Information about the consequences of a health problem helps people to recognize what could happen to them, whereas information about the antecedents of a health problem helps people to understand how it could happen. Because people rely on these two dimensions (i.e., cause and consequences) to structure their mental models of health problems *(12)*, information about the antecedents and consequences of a health problem has considerable informational value and can be readily utilized.

When people receive information about the consequences of a health problem, the implications of a potential health risk are rendered vivid and concrete. Although people may more readily recognize the severity of a potential health problem and may even be motivated to take precautionary action, imagining what it would be like to have a health problem does not provide people with sufficient information about how to address the potential problem. In the absence of information about how to effectively deal with a health problem, information about consequences may elicit greater feelings of personal risk and concern but, at the same time, may cause people to feel less confident about their ability to deal with a potential health threat. Because people need to know what they want to avoid and how they can go about minimizing their risk, merely drawing people's attention to the undesired consequences of a health problem is not likely to be an effective way to elicit risk-reducing behavior *(18,75)*.

Information about the antecedents of a health problem directs people's attention toward the factors that determine whether or not the problem will develop. Because antecedent information informs people about the specific factors that cause one's risk to increase, it can simultaneously provide them with information about what, if anything, they can do to minimize their risk. Thus, antecedent-based risk interventions offer the possibility of increasing not only people's perceptions of personal risk but also their confidence that they can cope with a potential health threat *(18,75)*. However, in and of itself, information about what causes a health problem to develop provides people with little information regarding the severity of the potential problem. People may learn how to minimize their risk, but in the absence of information about the severity of the problem they may not be sufficiently motivated to take action. Given that information about the consequences of a health problem provides people with a clear reason for taking action, the integration of information about the antecedents and the consequences of a health problem may prove to be an effective way to maximize the impact of a health risk intervention. Studies are needed that systematically examine the effect that the proposed integrative framework has on people's affective and behavioral responses to a health risk communication, with specific attention paid to the factors that are thought to mediate the impact of each type of information.

The development of a more precise understanding of how people respond to a health risk message should also help investigators anticipate *a priori* what factors are likely to moderate the message's impact. Given the heterogeneity of the health risks people must deal with, it is unlikely that a single intervention approach will prove effective for all health problems. Only with the formulation of a conceptual model of risk communication will we be in a position to predict whether an effective intervention strategy can be transferred successfully from one health domain to another. In a similar manner, even though studies have revealed that individual differences in personality and in prior experience can influence how people process health risk information [e.g, *(15,49)*], the extent to which these moderating factors generalize across health domains or interact with other aspects of the intervention strategy is not yet known (but *see* Gerrard et al., this monograph).

Can the Informational Value of Probability Information Be Increased?

Contextually based interventions have proven capable of influencing perceptions of risk, but these methodologies lack the precision afforded by a numerical estimate of the probability that an event will occur. Could an intervention integrate the presentation of probability information with information about the causes and consequences of a health problem? This approach would offer people precise information about the magnitude of their risk, but it would do so in a context that clarifies the implications of this information for their health. However, any intervention strategy that depends on the precision afforded by numerical probability information must confront the difficulties people have when forced to use this information *(39)*. If interventions are going to rely on people's ability to use probability information, then investigators need to provide them with the skills required to use the information accurately and appropriately. People need to know not only how to interpret numerical information about a specific health risk, but also how to gauge the degree to which adopting a particular precautionary behavior will affect these probabilities. Although there is considerable evidence that people can be taught to use basic statistical principles [e.g., *(76,77)*], the feasibility of integrating basic lessons in numeracy into interventions that provide people with health risk information remains untested.

Final Thoughts

If we expect people to make informed decisions about their health, information must be communicated to them in a manner that they can understand and use. Although scientific advances continue to provide increasingly more precise information about the health risks that people face, there has been little consensus as to the most effective way to communicate this information. On the basis of our review of the literature, we believe that intervention approaches that help people understand how a health problem could develop (i.e., its antecedents) and recognize what could happen to them (i.e., its consequences) offer the most effective way to communicate health risk information. However, the strength of this assessment is tempered by the absence of studies that have systematically compared the merits of different intervention approaches. A new generation of intervention studies is needed that will allow investigators to assess directly the relative impact of different risk communication strategies. However, the identification of effective communication strategies is not sufficient. Investigators need to determine not only what strategies are effective but also why they are effective. Substantial progress in our ability to transfer successful interventions across both different health domains and different participant populations will come only after we have developed a more precise understanding of how people process and utilize health risk information.

References

(1) Salovey P, Rothman AJ, Rodin J. Health behavior. In: Gilbert DT, Fiske ST, Lindzey G, editors. The handbook of social psychology, 4th ed. Vol. 2. Boston (MA): McGraw-Hill; 1998; p. 633–83.

(2) Weinstein ND. Testing four competing theories of health-protective behavior. Health Psychol 1993;12:324–33.

(3) Weinstein ND. Accuracy of smokers' risk perceptions. Ann Behav Med 1998;20:1–8.

(4) Kahneman D, Slovic P, Tversky A. Judgment under uncertainty: heuristics and biases. Cambridge (U.K.): Cambridge University Press; 1982.

(5) Taylor SE, Brown JD. Illusion and well-being: a social psychological perspective on mental health. Psychol Bull 1988;103:193–210.

(6) Ditto PH, Lopez DF. Motivated skepticism: use of differential decision criteria for preferred and nonpreferred conclusions. J Personality Social Psychol 1992;63:568–84.

(7) Gerrard M, Gibbons FX, Benthin AC, Hessling RM. A longitudinal study of the reciprocal nature of risk behaviors and cognitions in adolescents: what you do shapes what you think, and vice versa. Health Psychol 1996; 16:344–54.

(8) Liberman A, Chaiken S. Defensive processing of personally relevant health messages. Personality Social Psychol Bull 1992;18:669–79.

(9) Weinstein ND, Rothman AJ, Nicolich M. Using correlational data to examine the effects of risk perceptions on precautionary behavior. Psychol Health 1998;13:479–501.

(10) Slovic P. Perception of risk. Science 1987;236:280–5.

(11) Slovic P, Fischhoff B, Lichtenstein S. Facts versus fears: understanding perceived risk. In: Kahneman D, Slovic P, Tversky A, editors. Judgment under uncertainty: heuristics and biases. Cambridge (U.K.): Cambridge University Press; 1982; p. 463–89.

(12) Leventhal H, Nerenz DR, Steele DJ. Illness representations and coping with health threats. In: Baum A, Taylor SE, Singer J, editors. Handbook of psychology and health. Vol 4. Hillsdale (NJ): Erlbaum; 1984. p. 219–52.

(13) Armor DA, Taylor SE. Situated optimism: specific outcome expectancies and self-regulation. Adv Exp Social Psychol 1998;30:309–79.

(14) Kahneman D, Tversky A. The simulation heuristic. In: Kahneman D, Slovic P, Tversky A, editors. Judgment under uncertainty: heuristics and biases. Cambridge (U.K.): Cambridge University Press; 1982. p. 201–8.

(15) Rothman AJ, Schwarz N. Constructing perceptions of vulnerability: personal relevance and the use of experiential information in health judgments. Personality Social Psychol Bull 1998;24:1053–64.

(16) Kreuter MW, Strecher VJ. Changing inaccurate perceptions of health risk: results from a randomized trial. Health Psychol 1995;14:56–63.

(17) Janis IL. Effects of fear arousal on attitude change: recent developments in theory and experimental research. Adv Exp Social Psychol 1967;3: 166–224.

(18) Leventhal H. Findings and theory in the study of fear communications. Adv Exp Social Psychol 1970;5:119–86.

(19) Rothman AJ, Kelly KM, Weinstein ND, O'Leary A. Increasing the salience of risky sexual behavior: promoting HIV-antibody testing among heterosexually active young adults. J Applied Social Psychol 1999;29:531–51.

(20) Ajzen I. Intuitive theories of events and the effects of base-rate information on prediction. J Personality Social Psychol 1977;35:303–14.

(21) Tversky A, Kahneman D. Causal schemas in judgments under uncertainty. In: Kahneman D, Slovic P, Tversky A, editors. Judgment under uncertainty: heuristics and biases. Cambridge (U.K.): Cambridge University Press; 1982. p. 117–28.

(22) Klein WM, Kunda Z. Maintaining self-serving social comparisons: biased reconstruction of one's past behaviors. Personality Social Psychol Bull 1993;19:732–9.

(23) Rothman AJ, Klein WM, Weinstein ND. Absolute and relative biases about personal risk. J Applied Social Psychol 1996;26:1213–36.

(24) Klein WM, Weinstein ND. Social comparison and unrealistic optimism about personal risk. In: Buunk BP, Gibbons FX, editors. Health, coping, and well-being: perspectives from social comparison theory. Mahwah (NJ): Erlbaum; 1997. p. 25–61.

(25) Weinstein ND, Kolb K, Goldstein BD. Using time intervals between expected events to communicate risk magnitudes. Risk Analysis 1996;16: 305–8.

(26) Sandman PM, Weinstein ND, Miller P. High risk or low: how location on a "risk ladder" affects perceived risk. Risk Analysis 1994;14:35–45.

(27) Evers KE, Bishop CH, Gerhan LS, Weisse CS. AIDS educator effectiveness as a function of sexual orientation and HIV status. J Applied Social Psychol 1997;27:902–14.

(28) Gump BB, Kulik JA. The effect of a model's HIV status of self-perceptions: a self-protective similarity bias. Personality Social Psychol Bull 1995;21:827–33.

(29) Sutton SR, Hallett R. Understanding the effects of fear-arousing communications: the role of cognitive factors and amount of fear aroused. J Behav Med 1988;11:353–60.

(30) Sutton SR, Hallett R. Understanding seat-belt intentions and behavior: a decision-making approach. J Applied Social Psychol 1989;19:1310–25.

(31) Halpern DF, Blackman S, Salzman B. Using statistical information to assess oral contraceptive safety. Applied Cognitive Psychol 1989;3: 251–60.

(32) Harding CM, Eiser JR, Kristiansen CM. The representation of mortality statistics and the perceived importance of causes of death. J Applied Social Psychol 1982;12:169–81.

(33) Lippman-Hand A, Fraser FC. Genetic counseling: provision and reception of information. Am J Med Genet 1979;3:113–27.

(34) Parsons E, Atkinson P. Lay constructions of genetic risk. Sociol Health Illness 1992;14:437–55.

(35) Silka L, Albright L. Intuitive judgments of rate change: the case of teenage pregnancies. Basic Applied Social Psychol 1983;4:337–52.

(36) Black WC, Nease RF, Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst 1995;87:720–31.

(37) Doyle JK. Judging cumulative risk. J Applied Social Psychol 1997;27: 500–24.

(38) Shaklee H, Fischhoff B. The psychology of contraceptive surprises: cumulative risk and contraceptive effectiveness. J Applied Social Psychol 1990; 20:385–403.

(39) Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med 1997;127:966–72.

(40) Denes-Raj V, Epstein S, Cole J. The generality of the ratio-bias phenomenon. Personality Social Psychol Bull 1995;21:1083–92.

(41) Miller DT, Turnbull W, McFarland C. When a coincidence is suspicious: the role of mental simulation. J Personality Social Psychol 1989;57:581–9.

(42) Yamagishi K. When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. Applied Cognitive Psychol 1997;11: 495–506.

(43) Bottorff JL, Ratner PA, Johnson JL, Lovato CY, Joab SA. Communicating cancer risk information: the challenges of uncertainty. Patient Educ Counseling 1998;33:67–81.

(44) Palmer CG, Sainfort F. Toward a new conceptualization and operationalization of risk perception within the genetic counseling domain. J Genet Counseling 1993;2:275–94.

(45) Windschitl PD, Wells GL. Measuring psychological uncertainty: verbal versus numeric methods. J Exp Psychol Applied 1996;2:343–64.

(46) Roth E, Morgan MG, Fischhoff B, Lave L, Bostrom A. What do we know about making risk comparisons? Risk Analysis 1990;10:375–87.

(47) Klein WM. Objective standards are not enough: affective, self-evaluative, and behavioral responses to social comparison information. J Personality Social Psychol 1997;72:763–74.

(48) Blanton H, Gerrard M. Effect of sexual motivation on men's risk perception for sexually transmitted disease: there must be 50 ways to justify your lover. Health Psychol 1997;16:374–9.

(49) Gerrard M, Luus CA. Judgments of vulnerability to pregnancy: the role of risk factors and individual differences. Personality Social Psychol Bull 1995;21:160–71.

(50) Greening L, Chandler CC. Why it can't happen to me: the base rate matters, but overestimating skill leads to understanding risk. J Applied Social Psychol 1997;27:760–80.

(51) Kelly PT. Informational needs of individuals and families with hereditary cancers. Semin Oncol Nursing 1992;8:288–92.

(52) Weinstein ND. Persuasive conversations between friends about health and safety precautions. Health Educ Q 1993;20:347–60.

(53) Gerrard M, Gibbons FX, Warner TD. Effects of reviewing risk-relevant behavior on perceived vulnerability among women marines. Health Psychol 1991;10:173–9.

(54) Weinstein ND, Klein WM. Resistance of personal risk perceptions to debiasing interventions. Health Psychol 1995;14:132–40.
(55) Schwarz N, Bless H, Strack F, Klumpp G, Rittenauer-Schatka H, Simons A. Ease of retrieval as information: another look at the availability heuristic. J Personality Social Psychol 1991;61:195–202.
(56) Raghubir P, Menon G. AIDS and me, never the twain shall meet: the effects of information accessibility on judgments of risk and advertising effectiveness. J Consumer Res 1998;25:52–63.
(57) Sherman SJ, Cialdini RB, Schwartzman DF, Reynolds KD. Imagining can heighten or lower the perceived likelihood of contracting a disease: the mediating effect of ease of imagery. Personality Social Psychol Bull 1985; 11:118–27.
(58) Stapel DA, Reicher SD, Spears R. Social identity, availability and the perception of risk. Social Cognition 1994;12:1–17.
(59) Sutton SR, Eiser JR. The effect of fear-arousing communications on cigarette smoking: an expectancy-value approach. J Behav Med 1984;7:13–33.
(60) Sutton SR. Fear-arousing communications: a critical examination of theory and research. In: Eiser JR, editor. Social psychology and behavioral medicine. New York (NY): John Wiley & Sons; 1982. p. 303–37.
(61) Kong A, Barnett GO, Mosteller F, Yountz C. How medical professionals evaluate expressions of probability. N Engl J Med 1986;315:740–4.
(62) Mazur DJ, Hickman DH. Patients' interpretations of probability terms. J Gen Intern Med 1991;6:237–40.
(63) Borgida E, Nisbett RE. The differential impact of abstract vs. concrete information on decisions. J Applied Social Psychol 1977;7:258–71.
(64) Rook KS. Effects of case history versus abstract information on health attitudes and behaviors. J Applied Social Psychol 1987;17:533–53.
(65) Weinstein ND. Unrealistic optimism about future life events. J Personality Social Psychol 1980;39:806–20.
(66) Kiviniemi MT, Rothman AJ. Selective memory for information about health: an attitude congeniality effect. Poster session presented at the annual meeting of the American Psychological Association, 1998 Aug; San Francisco (CA).
(67) Smith GE, Gerrard M, Gibbons FX. Self-esteem and the relation between risk behavior and perceptions of vulnerability to unplanned pregnancy in college women. Health Psychol 1997;16:137–46.
(68) Kunda Z. The case for motivated reasoning. Psychol Bull 1990;108: 480–98.
(69) Croyle RT, Sande GN. Denial and confirmatory search: paradoxical consequences of medical diagnosis. J Applied Social Psychol 1988;18:473–90.
(70) Ditto PH, Jemmott JB III, Darley JM. Appraising the threat of illness: a mental representational approach. Health Psychol 1988;7:103–201.
(71) Croyle RT, Hunt JR. Coping with health threat: social influence processes in reactions to medical test results. J Personality Social Psychol 1991;60: 382–9.
(72) Croyle RT, Sun YC, Louie DH. Psychological minimization of cholesterol test results: moderators of appraisal in college students and community residents. Health Psychol 1993;12:503–7.
(73) McCaul KD, Thiese-Duffy E, Wilson P. Coping with medical diagnosis: the effects of at-risk versus disease labels over time. J Applied Social Psychol 1992;22:1340–55.
(74) McCaul KD, Schroeder DM, Reid PA. Breast cancer worry and screening: some prospective data. Health Psychol 1996;15:430–3.
(75) Taylor SE, Pham LB, Rivkin ID, Armor DA. Harnessing the imagination: mental simulation, self-regulation, and coping. Am Psychologist 1998;53: 429–39.
(76) Fong GT, Nisbett RE. Immediate and delayed transfer of training effects in statistical reasoning. J Exp Psychol General 1991;120:34–45.
(77) Nisbett RE, Fong GT, Lehman DR, Cheng PW. Teaching reasoning. Science 1987;238:625–31.

Notes

[1]Although this observation is frequently cited, no direct test of this claim could be identified. However, Halpern et al. *(31)* have found that people perceived oral contraceptives to pose a greater health risk when the information was presented as a relative risk (e.g., 415% greater risk of dying) than when it was presented as a base rate (e.g., one in 12 000 die).

[2]Although people primarily rely on the ease with which information comes to mind, they will base their judgment on the specific content of the recalled information when they are motivated to process the information in a more systematic and detailed manner *(15)* or when their subjective experience has been rendered nondiagnostic *(55)*.

[3]Of course, conflicting interpretations of information are not limited to numerical presentations of information. Several investigators have found there to be considerable variability in the values that people assign to quantitative labels such as frequently, sometimes, or occasionally [e.g., *(61,62)*].

Portions of this manuscript were presented at the Cancer Risk Communication Workshop organized by the National Cancer Institute.

We thank Rick Gibbons, Meg Gerrard, Steven Martino, and Neil Weinstein for their comments on an earlier version of this manuscript.

Cancer Screening Decisions

Kevin D. McCaul, Heather E. Tulloch

This review focuses on why people decide to obtain or to avoid screening for cancer. We discuss three topics: (*a*) physician prompts that may elicit compliant screening behavior, (*b*) the independent and joint effects of risk perceptions and worry, and (*c*) the costs and benefits of getting screened. Overall, the data suggest that each of these factors will influence screening. So, for example, people are more likely to seek screening if a physician recommends the behavior, if they feel personally vulnerable and worry a little about cancer, if insurance covers the screening, and if they believe that the test is an effective early detection procedure. Future research needs include studies comparing theories, longitudinal rather than cross-sectional studies, and true experiments. We also need to know more about why physicians are such powerful change agents and the trade-offs of increasing personal risk versus exacerbating worry. Practical recommendations for promoting cancer screening include encouraging physician interventions, explaining risk, and lowering the costs while emphasizing the benefits of screening. [Monogr Natl Cancer Inst 1999;25:52–8]

A variety of screening procedures, varying in effectiveness, is available for early detection of cancer. People also take advantage of such procedures to varying degrees *(1)*. Most women aged 18 years and over, for example, have had at least one Pap test—a remarkable adherence rate for a screening procedure. However, a much smaller proportion of Americans more than 50 years has had an examination for colon cancer, although this cancer can be cured if detected early *(2)*. How can we explain why people sometimes obtain and sometimes avoid cancer screening tests? The purpose of this review is to suggest empirical answers to this question, to propose directions for future research, and to address the implications of the literature for changing screening rates.

The literature review is organized into three sections. Within each section, we first look at correlational evidence and then, when possible, report experimental manipulations of the relevant variables. The three sections cover partially overlapping types of studies.

We propose that some cancer screening decisions may not be decisions in the typical sense of that word at all but rather may be more-or-less automatic responses to social influence. So, for example, when a 50-year-old man obtains a physical, he may participate in several screening procedures (an ear examination, a brief cardiogram, and a digital rectal examination) without actively deciding to do so. Instead, he simply might comply with the health professional's instructions. In the first section, we review the effects of a physician's recommendation on cancer screening rates.

The second section addresses two variables that theoretically set the stage for making more thoughtful, proactive decisions about screening: perceived risk of, and worry about, cancer. Most theories about health-protective behavior assume that people act to protect themselves against risk. Thus, people may decide to be screened because they feel vulnerable to cancer; in contrast, feelings of invulnerability should lead to reduced motivation to engage in cancer screening. Perceived risk is also related to worry about cancer, and each of these two variables may affect screening behavior.

Even if one feels vulnerable to cancer and worries about the disease, other psychosocial variables will influence the decision to be screened. These variables all fall under the general rubric of subjective utility theory *(3)*. This model suggests that people make rational decisions: They weigh the costs and benefits of a behavior before deciding whether to adopt it. Thus, people should see screening as valuable while carrying few costs if they are going to pursue a test.

The selection of variables comes from our analysis of the literature and existing theoretical approaches. In essence, we are proposing that these three conceptual categories will capture most of the important variance in predicting screening. This suggestion, however, could fairly be phrased as a hypothesis rather than a statement of fact. Future research may either attest to or disconfirm some of the conclusions that we draw.

The literature for the review was obtained with the use of two main search strategies. First, we searched the CD-ROM databases, PsychLIT® and MEDLINE®, from 1981 through 1998. The key terms used included "physician," "cancer screening," "breast cancer screening," "cervical cancer," "colorectal cancer," "predictors," "motivation," "skin cancer," "melanoma," "decisions," "perceived risk," "worry," "mammography," and "mammogram." This search process resulted in articles examining screening for six types of cancer: breast, colon, skin, cervical, prostate, and ovarian. Because we were interested in why people obtain cancer screening, we included articles for all of these types of cancers, although the different screening tools vary widely in their effectiveness *(4)*. The second approach was to examine the reference lists of the published articles found. Given the space limitations, we have not attempted to cite all articles about each conceptual issue. Instead, we have selected articles to illustrate the conceptual arguments and to suggest that the answers generalize across different cancer screening procedures.

Physician Recommendations

Most theories of health-protective behavior probably underestimate the effects of social influence. An exception would be the general model of volitional behavior offered by Fishbein and Ajzen *(5)*, the Theory of Reasoned Action (TRA). The TRA

Affiliation of authors: Department of Psychology, North Dakota State University, Fargo.

Correspondence to: Kevin D. McCaul, Ph.D., Department of Psychology, North Dakota State University, 201 Stevens Hall, Fargo, ND 58105 (e-mail: mccaul@badlands.nodak.edu).

See "Notes" following "References."

© Oxford University Press

states that normative beliefs (and the more general "subjective norm") is one of two immediate precursors of intentions to act. Normative beliefs include a person's evaluation of what significant others believe he or she should do and how motivated he or she is to comply with them. The TRA does not suggest *a priori* that one person should be more influential in affecting our cancer screening decisions than any other individual. However, although some data support the effects of people other than physicians in influencing screening decisions *(6,7)*, findings from many studies show the power of physicians to affect screening behavior. Thus, we have created a separate category for physician recommendation based more on empirical findings than because of any special theoretical significance.

Most research concerning physician recommendations (indeed, most research about cancer screening overall) comes from the breast cancer screening literature. Women who have had mammograms, for example, report that a physician recommendation motivated their behavior *(8)*. Women who have not had mammograms report the obverse: "My doctor did not tell me I needed to be screened" *(9,10)*. A physician recommendation correlates with both initial and repeat mammograms *(11)*, and the power of such a recommendation extends to other kinds of cancer screenings and beyond North America. For example, inner-city American women reported that they had not obtained a Pap smear because a doctor had not told them to do so [*(12)*; also, *see (13)*]. In London, U.K., a survey showed that physician practices in which staff contacted women to encourage screening had higher breast and cervical cancer screening rates *(14)*. Finally, the data suggest not only that a physician recommendation is correlated with screening but that it is also the single most powerful determinant of screening *(15)*.

Correlational data, of course, are only suggestive of cause and effect, and most of the data supporting the importance of physician recommendations are correlational. Still, some researchers have attempted to increase screening by changing physician behavior [*see (16)*]. A recent review of cervical cancer screening interventions noted that physician and patient prompts were successful in promoting Pap smears and that "opportunistic screening" intervention (same-day examinations during an outpatient or hospital visit) were especially effective *(17)*. Adding a computerized physician reminder increased mammography rates compared with an intervention that did not require direct physician participation *(18)*. A physician letter and phone call produced fourfold increases in cancer screening (Pap smear and mammography) compared with a no-treatment condition *(19)*. Although most experiments have compared physician interventions with control conditions with no additional interventions beyond usual care, a few experiments have compared physician interventions with more sophisticated alternative attempts to increase screening. Clover et al. *(20)*, for example, found that physician interventions produced higher mammography rates than did media or comprehensive community campaigns.

Risk and Worry

General theories of attitude–behavior linkages, such as the TRA, do not specifically address the importance of perceived risk [Stasson and Fishbein *(21)* even argued that differentiating perceived risk from other beliefs is unnecessary]. However, all theoretical models specifically addressing health-protective behaviors emphasize perceived risk *(22,23)*.[1] Risk is defined here as the likelihood that one might have cancer at some future date; investigators also use the term "susceptibility" and "vulnerability" to mean the same thing. Other elements of risk (e.g., novelty) could be related to screening *(24)*, but health-protective theories have focused on the likelihood dimension.

The relationship between risk and screening can be tested directly by correlating perceived risk with screening frequency. However, the relationship can also be inferred from correlations with two other "susceptibility" variables: whether one has a family history of cancer and whether one has experienced problems or symptoms related to cancer. Perceived cancer risk is also related, although imperfectly, to worry about cancer. Some theorists suggest that risk, a cognitive variable, and worry, an affective variable, may each be related to action, and we will examine that possibility. The remainder of this section reviews each of these four risk-related variables: perceived risk, family history, symptoms, and worry.

Perceived Risk

McCaul et al. *(25)* reported a meta-analysis relating perceived risk to mammography screening, obtaining a small but significant effect size of $r = .16$ from 19 studies. Perceived vulnerability is also positively correlated with screening for other cancers [e.g., colon cancer *(26)*, skin cancer *(27–29)*, and prostate cancer *(30)*]. However, perceived risk is but a single variable among many decision-making variables that could influence an infrequently performed behavior. Not surprisingly, then, one is likely to observe only modest effect sizes obtained for risk and screening behavior relationships. Given such effect sizes, it would also be expected that some individual studies will fail to produce statistically significant relationships *(25)*.

Because risk perceptions play a central role in theories of health-protective behavior, we were surprised to find no experiments in which researchers explicitly manipulated risk and examined screening outcomes. In two studies, however, investigators emphasized risk status to people with or without a family history of cancer. Blalock et al. *(31)* told siblings of persons diagnosed with colorectal cancer about their individual risk and urged them to do a home screening test. These persons were more likely to do the test than persons without a family history who received more general risk feedback. Family history was also an important background variable in a study reported by Curry et al. *(32)* concerning mammography screening. Women with a family history of breast cancer who received a mammography invitation that mentioned their personal risk were more likely to obtain a mammogram than were similar women who received an invitation that only addressed their risk in general terms.

Family History

An observed relationship between family history of cancer and a greater likelihood of screening would not necessarily be related to active decision making. A positive family history could influence screening through multiple routes: prompts from physicians, an increased sense of personal vulnerability, or increased worry. It is likely that physicians are more likely to recommend screening for women with a positive family history, but there are few data. However, there is evidence that a family history correlates with increases in perceived risk of breast cancer *(33)* and skin cancer *(34)*.[2] A family history of breast cancer is also related to increased worry about breast cancer *(35,36)*. McCaul et al. *(35)* asked women with and without a family

history of breast cancer to self-monitor their thoughts and worry for a month. Worry declined over time, perhaps because it was elevated initially in association with the context of the study. However, women with a family history of the disease showed a smaller decline in worry, and the researchers suggested that such women could be chronically worried about the disease.

Although the mechanism is unclear, a positive family history of cancer clearly predicts higher screening levels. McCaul et al. *(25)* found an effect size of $r = .27$ from 19 studies examining a link between family history of breast cancer and mammography screening. That effect size was meaningfully higher than the effect size for perceived risk alone ($r = .16$), which makes sense if a family history creates several possible routes to increased screening. Similar relationships have also been observed for cancers other than breast cancer. A positive family history of prostate cancer predicts interest in prostate-specific antigen (PSA) testing *(37)*, and a lack of a family history predicts poorer adherence to Pap smear screening *(38)*. A family history of breast cancer is also related to higher intentions to do breast self-examination (BSE) *(35)*. Richardson et al. *(39)* discovered that unaffected co-twins were screened for colorectal cancer at a rate three to four times higher than the general population. Sandler et al. *(40)* found that high-risk siblings of diagnosed colon cancer patients were more interested in fecal occult blood testing than controls.

Symptoms

Similar to a family history, a link between the presence of symptoms and screening is not necessarily related to active decision making about screening. Symptoms could prompt a physician recommendation *(41)*, increase perceived risk, or elevate worry.[3] We did not find data concerning the first two possibilities, but Cunningham et al. *(42)* found that women with benign breast problems reported more breast cancer worry than women without such problems.

Data show that symptoms predict higher screening rates. Women with breast problems, for example, are more likely to have had a mammogram *(43)*. McCaul et al. *(25)* obtained an effect size of .30 between the discovery of breast problems and having had mammograms. The presence of risk factors, including symptoms, was associated prospectively with colorectal cancer screening *(44)*. Cameron et al. *(45)* found that high-anxiety women on tamoxifen were more adherent to BSE than low-anxiety women. They attributed this finding to the possibility that high-anxiety women used the symptoms caused by tamoxifen as cues to prompt testing. One of the more frightening "symptoms" one could have is a false-positive mammogram that requires a surgical biopsy. One study has shown that women having such an experience may be more likely to obtain future mammograms as recommended by guidelines *(46)*.

Worry

Cancer is probably the most feared contemporary disease *(47)*, so it is not surprising that people worry about it. Some authors have argued that worry can impede screening *(48)*, but that suggestion conflicts with both theory and data. Theoretically, Leventhal *(49)* has most directly addressed the role of affect, suggesting that distress about a health threat will motivate self-protective behavior if an action exists that can reduce the threat. Thus, Leventhal would suggest that both risk and worry should contribute to higher screening levels [*see (50)*]. Leventhal's insertion of affect into what is typically seen as a more rational decision is important, and his parallel model fits with more general theorizing about the dual aspects of consciousness [cf. *(51)*].

McCaul et al. *(25)* observed a small but significant positive relationship ($r = .14$) between worry and mammography screening on the basis of only six studies. McCaul and colleagues have also found positive correlations between worry and higher levels of BSE. These relationships between worry and breast cancer screening hold both concurrently and prospectively *(35,52,53)*. We were unable to find many reports of connections between cancer worry and screening for cancers other than breast cancer. However, De Rooij et al. *(54)* recently reported data from The Netherlands that fear of having skin cancer was an important reason given for attending a screening clinic for melanoma screening.

One might speculate that worry is not the crucial variable predicting screening but serves as a proxy for perceived risk. Risk and worry are related, although modestly. Linville et al. *(55)* have shown that correlations between risk and worry average about .30 across many behaviors. McCaul and O'Donnell *(56)* found a correlation of $r = .36$ between worry, defined as intrusive thoughts, and breast cancer risk. Webb et al. *(34)* obtained a risk–worry correlation for skin cancer of $r = .19$, and Schwartz et al. *(57)* observed an $r = .25$ between risk and worry about ovarian cancer. It is possible that both risk and worry predict more frequent screening. Some research suggests that this is the case—partial correlations between worry and breast cancer screening stayed significant after partialing out perceived risk *(52,53)*. Schwartz et al. *(58)* also recently showed that family history (an indirect measure of perceived risk) and worry were both independent, positive predictors of screening for ovarian cancer.

Not surprisingly, researchers will not purposefully increase worry to determine whether increased negative affect increases cancer screening. However, some correlational data about the effects of mammography screening lend weight to the possibility that worry increases screening. Lerman et al. *(59)* interviewed women about their worry just after they had received a mammogram. Women who said that they were still worried about breast cancer also reported that they were more likely to obtain another mammogram. Similarly, in another study, Lerman et al. *(60)* reported that women who expressed relief after a mammogram were less likely to return for a subsequent screening.

The preponderance of data suggest that breast cancer worries do not lead to screening avoidance but just the opposite—worry leads people to try to protect themselves by taking advantage of screening opportunities. Some people, however, may use screening excessively. Two studies report data suggesting that affect can prompt excessive self-examination for breast cancer. Lerman et al. *(61)* discovered that young women who had both a family history of breast cancer and a high level of breast cancer anxiety were more likely to perform BSE more than once a month. More troublesome, perhaps, Epstein et al. *(62)* interviewed women who had recently had a first-degree relative diagnosed with a primary breast cancer. A few of those women (8%) reported that they examined their breasts for lumps every day. These women were more likely than the rest of the sample to report that thinking about breast cancer negatively affected their mood—one of three items that the authors used to measure worry. The women who were too frequently

examining their breasts were also more confident about protecting their health and about doing something about their cancer worries, suggesting that they perceived the constant breast checks positively.

Finally, we should note that differentiating between anxiety and worry about cancer screening could be important. Worry clearly can have positive effects. Most prominently, worry predicts problem solving; Davey *(63)* called it "constructive worry" [*see* also *(64)*]. Anxiety may not have such positive outcomes, and measures of anxiety have typically not been associated with higher screening levels. McCaul et al. *(35)*, for example, failed to observe a positive relationship between trait anxiety and screening. And, for women with a strong family history of breast cancer, Kash et al. *(65)* observed a negative relationship between anxiety and screening.

Costs and Benefits of Acting

Some version of the costs and benefits of adopting a health-protective behavior is a part of all models of health protection. The Health Belief Model *(66)*, for example, includes a single cost/benefit variable; Protection Motivation Theory *(67)* addresses the cost of performing and the likely effectiveness of the behavior. The "pros and cons" of adopting a health-protective behavior are also theoretically important in moving people to different stages in the Transtheoretical Model *(68)*. Costs and benefits are also important factors in more general models of attitude–behavior linkages (e.g., the TRA). We next consider costs (barriers) and benefits in turn.

Barriers

In a recent review, Womeodu and Bailey *(69)* listed 25 barriers to cancer screening. Some factors that inhibit screening for different screening tests include financial cost [ovarian *(70)* and breast *(71)*]; concern about radiation [breast *(72)*]; embarrassment [cervical *(73)*], ease of access, including travel difficulties [breast *(72)* and ovarian *(70)*], anxiety about test results [mammography *(74)* and BSE *(75)*], not having symptoms [cervical *(76)* and prostate *(77)*], inconvenience [colorectal *(78)*], forgetting or procrastination *(79)*, and discomfort associated with the screening test [prostate *(80)*].

Some researchers have tested interventions intended to reduce barriers, and these interventions are often successful. Wolosin *(81)*, for example, showed that forgetting can be overcome by scheduling specific appointments and reminding women. Reminders increased screening for several forms of coloreetal cancer screening *(82)*. Several studies have relied on tailoring the interventions to the particular barriers that women express. King et al. *(72)* doubled mammography screening rates by calling older women and counseling them how to overcome screening barriers. Skinner et al. *(83)* tailored mailed messages to the barriers to mammography that women had reported earlier; this strategy produced a 13% increment in subsequent screening compared with a nontailored message. More recently, Rakowski et al. *(84)* tailored a mailed message to participants' "stage of change." They reported a significant difference in subsequent mammogram rates of 64% versus 55% in a no-mailing condition.

Benefits

Benefits to cancer screening seem fewer than possible barriers. The overriding benefit may be whether the screening test truly confers the likelihood of early detection *(20,44)* or at least that persons believe the test to be valuable for early detection *(85)*. A related notion is whether the examination provides peace of mind *(86)*. We did not identify any interventions that focused solely on benefits, which makes sense—researchers would be likely to balance such an emphasis by also reducing barriers. But almost any intervention, regardless of its particular components, will at least include information on the benefits of screening.

Thus far in this review, we have ignored demographic variables that could be associated with screening. We speculate that the effects of such variables are usually mediated through the sorts of psychosocial variables already covered. If, for example, income relates positively to screening levels, any of the following (among other) mediated relationships could explain the connection: (*a*) people with high income may visit physicians more frequently and thus receive screening recommendations; (*b*) people with higher income are also more educated, and they could be more knowledgeable about their risk; and (*c*) people with higher income are more likely to have insurance coverage, reducing the cost barrier. Other variables, such as age, may have their effects on screening levels through similarly mediated relationships.

One exception to the above could be ethnicity. Women from different cultural backgrounds may behave differently in terms of screening because of cultural beliefs *(87,88)*. Hubbell et al. *(89)* discovered that cultural beliefs concerning risk factors might inhibit Latina immigrants from obtaining Pap smears.

Research Needs

We suggest that cross-sectional studies correlating various demographic or psychosocial factors with screening behaviors are of less value at this time than other types of research. Studies that produce empirically, rather than theoretically, derived regression equations predicting screening are also not a priority. Two other kinds of correlational studies would prove more useful. First, theoretically driven studies that compare theories would help inform interventions *(23)*. Second, such studies would be more valuable if researchers followed participants over time. Finally, we need research in which investigators manipulate variables in true experiments.

What is important to discover that we do not yet know? One possibility is to conduct further research to understand the relationship between physician recommendations and screening behavior. Why are physicians such powerful social influence agents? Is it merely obedience to authority? Is it because physicians are especially good at conveying a sense of personal vulnerability or worry? We do not know the answers to these kinds of questions. A second possible direction is to test the speculation in this paper concerning the effects of worry versus perceived risk. Does worry have different effects in predicting screening than other types of affective responses such as anxiety?

Practice Implications

If one wished to encourage screening, what could be done? Before considering plausible answers to this question, we should acknowledge that the goal may not always be to encourage people to be screened. Instead, it may be only to provide information so that people can decide for themselves. The latter goal should probably be primary when the value of screening is unknown, but some authors also argue that such a strategy is al-

ways the appropriate choice *(90,91)*. Of course, when information alone is provided, the effects on screening may be to either increase or decrease rates. Wolf and Schorling *(30)*, for example, showed that providing accurate information about the efficacy of the PSA test to detect prostate cancer reduced interest in obtaining the examination among elderly men. Of course, this is a cancer whose role in reducing mortality has not been proven.

To the extent that our goal is to increase screening, the three types of variables we have reviewed lead to obvious practical implications. First, health professionals should simply recommend to their patients that they obtain a screening, scheduling it at the same time as the recommendation, if possible. Do physicians do so? Lerman et al. *(11)* reported that only 60% of their respondents said that their physicians had recommended mammograms. These data come from early 1988, however, and it is likely that a greater proportion of physicians would recommend mammographies today. Other cancer screening tests are probably recommended less often. Demark-Wahnefried et al. *(86)*, for example, discovered that only 28% of black or white men reported that their doctor ever discussed a prostate cancer test with them.

Although a physician recommendation will promote screening, some people still may not be screened for at least two reasons. First, of course, everyone does not visit a physician regularly; we must find other means to reach such individuals. Second, although we have described compliance with a physician recommendation as more or less "automatic," that description is not always accurate, especially because time often intervenes between the recommendation and actually scheduling or obtaining a screening. As the time interval following a recommendation increases, more active decision making may be important. Thus, a second recommendation would be to give people information showing that they are personally vulnerable to the cancer associated with the screening.

The recommendation to emphasize personal risk runs head on into possible ethical issues. Ensuring that accurate information is presented will occasionally mean convincing people that their risk is higher than they believe. The analysis presented earlier suggests that, if persuasion is effective, people will raise their risk estimates, worry more about cancer, and be more likely to adopt a screening behavior. Thus, this strategy has costs (occasional worry). We were unable to find any studies in the literature that addressed the use of risk counseling to increase risk estimates.

A more likely possibility is that counseling about accuracy will mean convincing many people that their risk is lower than they believe. This is because many people overestimate their risk, at least when absolute (rather than social comparative) measures of risk are used *(56)*. Reducing risk estimates may have ironic costs according to our analysis—a lower perceived risk of screening accompanied by less worry but also reduced motivation to be screened. This would not necessarily have to be the case, however, because risk and worry—although important—are only two of many variables influencing screening. Counseling could effectively shift screening control from individual decision making to professional "control" (e.g., reminders from one's physician). Counseling could also focus on removing the barriers that inhibit screening (especially concern about the screening procedure per se) and influencing beliefs about the value of screening. Thus, although lower risk estimates should reduce screening if all other things are equal, the latter caveat is unlikely to hold for most counseling programs.

Preliminary data concerning the effects of providing accurate information have been interesting. Women who are overestimating their risk reduce their estimates, but not as low as the levels provided by the risk counselors *(92)*. The only study *(93)* to measure actual screening behavior showed that a risk counseling program for women with a family history of breast cancer lowered risk estimates and distress, but reduced the number of women who had a mammogram during the subsequent year. Other studies *(94,95)*, however, show that women report less distress about cancer but are no less motivated to engage in screening.

Finally, if we wish to increase screening levels, we can remove barriers and increase beliefs in the value of screening. Many interventions rely on this general strategy *(96,97)*. Thus, for example, to the extent that screening has been shown to have a public health benefit, we can work to reduce costs, ease access, and improve screening technology. For effective screening techniques, all such attempts to improve the benefit/cost ratio will likely increase screening rates.

References

(1) Vernon SW. Risk perception and risk communication for cancer screening behaviors: a review. Monogr Natl Cancer Inst 1999;25:101–19.

(2) Bowdy M. Lag in colorectal screening rates prompts innovation. J Natl Cancer Inst 1998;90:886–7.

(3) Edwards W. The theory of decision making. Psych Bull 1954;51:380–417.

(4) U.S. Preventive Services Task Force. Guide to clinical preventive services: report of the U.S. Preventive Services Task Force. 2nd ed. Baltimore (MD): Williams & Wilkins; 1996.

(5) Fishbein M, Ajzen I. Belief, attitude, intention, and behavior: an introduction to theory and research. Reading (MA): Addison-Wesley; 1975.

(6) Allen JD, Sorensen G, Stoddard AM, Colditz G, Peterson K. Intention to have a mammogram in the future among women who have underused mammography in the past. Health Ed Behav 1998;25:474–88.

(7) McCance KL, Mooneh KH, Field R, Smith KR. Influence of others in motivating women to obtain breast cancer screening. Cancer Prac 1996;4: 141–6.

(8) Fox S, Murata PJ, Stein JA. The impact of physician compliance on screening mammography for older women. Arch Intern Med 1991;151:50–6.

(9) Frazier EL, Jiles RB, Mayberry R. Use of screening mammography and clinical breast examinations among Black, Hispanic, and White women. Prev Med 1996;25:118–25.

(10) O'Connor AM, Perrault DJ. Importance of physician's role highlighted in survey of women's breast screening practices. Can J Publ Health 1995;86: 42–5.

(11) Lerman C, Rimer B, Trock B, Balshem A, Engstrom PF. Factors associated with repeat adherence to breast cancer screening. Prev Med 1990;19: 279–90.

(12) Mamon JA, Shediac MC, Crosby CB, Sanders B, Matanoski GM, Celentano D. Inner-city women at risk for cervical cancer: behavioral and utilization factors related to inadequate screening. Prev Med 1990;19:363–76.

(13) Harlan LC, Bernstein AB, Kessler LG. Cervical cancer screening: who is not screened and why? Am J Publ Health 1991;81:885–90.

(14) Majeed FA, Cook D, Given-Wilson R, Vecchi P, Poloniecki J. Do general practitioners influence the uptake of breast cancer screening? J Med Screening 1995;2:119–24.

(15) Rimer BK. Understanding the acceptance of mammography by women. Ann Behav Med 1992;14:197–203.

(16) Snell JL, Buck EL. Increasing cancer screening: a meta-analysis. Prev Med 1996;25:702–7.

(17) Marcus AC, Crane LA. A review of cervical cancer screening intervention research: implications for public health programs and future research. Prev Med 1998;27:13–31.

(18) Burack R, Gimotty P, George J, Stingle W, Warbasse L, Moncrease A

Promoting screening mammography in inner-city settings: a randomized controlled trial of computerized reminders as a component of a program to facilitate mammography. Med Care 1994;32:609–24.
Lantz PM, Stencil D, Lippert M, Beversdorf S, Jaros L, Remington P. Breast and cervical cancer screening in a low-income managed care sample: the efficacy of physician letters and phone calls. Am J Publ Health 1995;85:834–6.
Clover K, Redman S, Forbes J, Sanson-Fisher R, Calleghan T. Two sequential randomized trials of community participation to recruit women for mammographic screening. Prev Med 1996;25:126–34.
Stasson M, Fishbein M. The relationship between perceived risk and preventive action: a within-subject analysis of perceived driving risk and intentions to wear seatbelts. J Appl Soc Psychol 1990;20:1541–57.
Van der Pligt J. Risk appraisal and health behavior. In: Rutter DR, Quine L, editors. Social psychology and health: European perspectives. Aldershot (U.K.): Avebury; 1994. p. 131–51.
Weinstein ND. Testing four competing theories of health-protective behavior. Health Psychol 1993;12:324–33.
Fischoff B. Worry: what it gets you? Paper presented at the conference on Worry; 1991 Jan; San Diego (CA).
McCaul KD, Branstetter AD, Schroeder DM, Glasgow RE. What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. Health Psychol 1996;15:1–8.
Macrae FA, Hill DJ, St. John DJ, Ambikapathy A, Garner JF, the Ballarat General Practitioner Research Group. Predicting colon cancer screening behavior from health beliefs. Prev Med 1984;13:115–26.
Cody R, Lee C. Behaviors, beliefs, and intentions in skin cancer prevention. J Behav Med 1990;13:373–89.
Friedman LC, Webb JA, Bruce S, Weinberg AD, Cooper HP. Skin cancer prevention and early detection intentions and behavior. Am J Prev Med 1995;11:59–65.
Girgis A, Campbell EM, Redman S, Sanson-Fisher RW. Screening for melanoma: a community survey of prevalence and predictors. Med J Aust 1991;154:338–43.
Wolf AMD, Schorling JB. Preferences of elderly men for prostate-specific antigen screening and the impact of informed consent. J Geront Med Sci 1998;53:M195–200.
Blalock SJ, DeVellis BM, Afifi RA, Sandler RS. Risk perceptions and participation in colorectal cancer screening. Health Psychol 1990;9: 792–806.
Curry SJ, Taplin SH, Anderman C, Barlow WE, McBride C. Randomized trial of the impact of risk assessment and feedback on participation in mammography screening. Prev Med 1993;22:350–60.
Drossaert CC, Boer H, Seydel ER. Perceived risk, anxiety, mammogram uptake, and breast self-examination of women with a family of breast cancer: the role of knowing to be at increased risk. Cancer Detect Prev 1996;20:76–85.
Webb JA, Friedman LC, Bruce SB, Weinberg AD, Cooper P. Demographic, psychosocial, and objective risk factors related to perceived risk of skin cancer. J Cancer Educ 1996;11:174–7.
McCaul KD, Branstetter AD, O'Donnell SM, Jacobson K, Quinlan KB. A descriptive study of breast cancer worry. J Behav Med 1998;21:565–79.
Valdimarsdottir HB, Bovbjerg DH, Kash KM, Holland JC, Osborne, MP, Miller DG. Psychological distress in women with a familial risk of breast cancer. Psycho-oncology 1995;4:133–41.
Wolf AM, Philbrick JT, Schorling JB. Predictors of interest in prostate-specific antigen screening and the impact of informed consent: what should we tell our patients? Am J Med 1997;103:308–14.
Weinrich S, Coker AL, Weinrich M, Eleazer GP, Greene FL. Predictors of pap smear screening in socioeconomically disadvantaged elderly women. J Am Geriatr Soc 1995;43:267–70.
Richardson JL, Danley K, Mondrus GT, Deapen D, Mack T. Adherence to screening examinations for colorectal cancer after diagnosis in a first-degree relative. Prev Med 1995;24:166–70.
Sandler RS, DeVellis BM, Blalock SJ, Holland KL. Participation of high-risk subjects in colon cancer screening. Cancer 1989;63:2211–5.
Grady KE, Lemkau JP, McVay JM, Reisine ST. The importance of physician encouragement in breast cancer screening of older women. Prev Med 1992;21:766–80.
Cunningham LL, Andrykowski MA, Wilson JF, McGrath PC, Sloan DA, Kenady DE. Physical symptoms, distress, and breast cancer risk perceptions in women with benign breast problems. Health Psychol 1998;17: 371–5.
(43) Zapka JG, Costanza ME, Greene HL Jr. Breast cancer screening and primary care providers. Prog Clin Biol Res 1989;293:55–63.
(44) Myers RE, Trock BJ, Lerman C, Wolf T, Ross E, Engstrom PF. Adherence to colorectal cancer screening in an HMO population. Prev Med 1990;19: 502–14.
(45) Cameron LD, Leventhal H, Love RR. Trait anxiety, symptom perceptions, and illness-related responses among women with breast cancer in remission during a tamoxifen clinical trial. Health Psychol 1998;17:459–69.
(46) Pisano ED, Earp J, Schell M, Vokaty K, Denham A. Screening behavior of women after a false-positive mammogram. Radiology 1998;208:245–9.
(47) Murray M, McMillan CL. Gender differences in perceptions of cancer. J Cancer Educ 1993;8:53–62.
(48) Hailey BJ. Family history of breast cancer and screening behavior: an inverted U-shaped curve? Med Hypotheses 1991;36:397–403.
(49) Leventhal H. Findings and theory in the study of fear communications. In: Berkowitz L, editor. Advances in experimental social psychology. Vol 3. San Diego (CA): Academic Press; 1970. p. 110–86.
(50) Sutton SR, Eiser JR. The decision to wear a seat belt: the role of cognitive factors, fear and prior behaviour. Psychol Health 1990;4:111–23.
(51) Epstein S. Integration of the cognitive and the psychodynamic unconscious. Am Psychol 1994;49:709–24.
(52) McCaul KD, Reid PA, Rathge RW, Martinson B. Does concern about breast cancer inhibit or promote breast cancer screening? Basic Appl Socl Psych 1996;18:183–94.
(53) McCaul KD, Schroeder DM, Reid PA. Breast cancer worry and screening: some prospective data. Health Psychol 1996;15:430–3.
(54) De Rooij MJ, Rampen FH, Schouten LJ, Neumann HA. Factors influencing participation among melanoma screening attenders. Acta Derm Venereol 1997;77:467–70.
(55) Linville PW, Fishcher GW, Fischhoff B. AIDS risk perceptions and decisions biases. In: Pryor JB, Reeder GD, editors. The social psychology of HIV infection. Hillsdale (NJ): Erlbaum; 1993. p. 5–38.
(56) McCaul KD, O'Donnell SM. Naive beliefs about breast cancer risk. Womens Health 1998;4:93–101.
(57) Schwartz MC, Lerman C, Miller SM, Daly M, Masny A. Coping disposition, perceived risk, and psychological distress among women at increased risk for ovarian cancer. Health Psychol 1995;14:232–5.
(58) Schwartz M, Lerman C, Daly M, Audrain J, Masny A, Griffith K. Utilization of ovarian cancer screening by women at increased risk. Cancer Epid Biomarkers Prev 1995;4:269–73.
(59) Lerman C, Trock B, Rimer B, Boyce A, Jepson C, Engstrom P. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med 1991;114:657–61.
(60) Lerman C, Trock B, Rimer B, Jepson C, Brody D, Boyce A. Psychological side effects of breast cancer screening. Health Psychol 1991;10:259–67.
(61) Lerman C, Kash K, Stenfanek M. Younger women at increased risk for breast cancer: perceived risk, psychological well-being, and surveillance behavior. Monogr Natl Cancer Inst 1994;16:171–6.
(62) Epstein SA, Lin TH, Audrain J, Stefanek M, Rimer B, Lerman C. Excessive breast self-examination among first-degree relatives of newly diagnosed breast cancer patients. Psychosomat 1997;38:253–61.
(63) Davey GC. A comparison of three worry questionnaires. Behav Res Ther 1993;31:51–6.
(64) Fischoff B, Bostrom A, Quadrel MJ. Risk perception and communication. Annu Rev Publ Health 1993;14:183–203.
(65) Kash KM, Holland JC, Halper MS, Miller DG. Psychological distress and surveillance behaviors of women with a family history of breast cancer. J Natl Cancer Inst 1992;84:24–30.
(66) Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11:1–47.
(67) Rogers RW. A protection motivation theory of fear appeals and attitude change. J Psych 1975;91:93–114.
(68) Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promotion 1997;12:38–48.
(69) Womeodu RJ, Bailey JE. Barriers to cancer screening. Med Clin North Am 1996;80:115–33.
(70) Pavlik EJ, Van Nagell JR, DePriest PD, Wheeler L, Tatman JM, Boone M,

et al. Participation in transvaginal ovarian cancer screening: compliance, correlation factors, and costs. Gynecol Oncol 1995;57:395–400.
(71) Rutledge DN, Hartmann WH, Kinman PO, Winfield AC. Exploration of factors affecting mammography behaviors. Prev Med 1988;17:412–22.
(72) King ES, Balshem A, Ross E, Rimer B, Seay J. Mammography interventions for 65- to 74-year-old HMO women. J Aging Health 1995;7:529–51.
(73) Peters RK, Bear MB, Thomas D. Barriers to screening for cancer of the cervix. Prev Med 1989;18:133–46.
(74) AMC Cancer Research Center. Breast and cervical cancer screening: barriers and use among specific populations. Atlanta (GA): U.S. Centers for Disease Control and Prevention; 1995.
(75) Murray M, McMillan C. Social and behavioural predictors of women's cancer screening practices in Northern Ireland. J Publ Health Med 1993; 15:147–53.
(76) McKie L. Women's views of the cervical smear test: implications for nursing practice—women who have not had a smear test. J Adv Nurs 1993;18:1228–34.
(77) Nijs HG, Tordoir DM, Schuurman JH, Kirkels WJ, Schroder FH. Randomised trial of prostate cancer screening in The Netherlands: assessment of acceptance and motives for attendance. J Med Screening 1997;4:102–6.
(78) Myers RE, Ross E, Jepson C, Wolf T, Balshem A, Millner L, et al. Modeling adherence to colorectal cancer screening. Prev Med 1994;23:142–51.
(79) Farrands PA, Hardcastle JD, Chamberlain J, Moss S. Factors affecting compliance with screening for colorectal cancer. Comm Med 1984;6:12–9.
(80) McCoy CB, Anwyl RS, Metsch LR, Inciardi JA, Smith SA, Correa R. Prostate cancer in Florida: knowledge, attitudes, practices, and beliefs. Cancer Practice 1995;3:88–93.
(81) Wolosin RJ. Effect of appointment scheduling and reminder postcards on adherence to mammography recommendations. J Fam Pract 1990;30: 542–7.
(82) Vernon SW. Participation in colorectal cancer screening: a review. J Natl Cancer Inst 1997;89:1406–22.
(83) Skinner CS, Strecher VJ, Hospers H. Physicians' recommendations for mammography: do tailored messages make a difference? Am J Publ Health 1994;84:43–9.
(84) Rakowski W, Ehrich B, Goldstein M, Rimer BK, Pearlman DN, Clark MA, et al. Increasing mammography among women aged 40–74 by use of a stage-matched, tailored intervention. Prev Med 1998;27:748–56.
(85) Tingen MS, Weinrich SP, Boyd MD, Weinrich MC. Prostate cancer screening: predictors of participation. J Am Acad Nurs Prac 1997;9:557–67.
(86) Demark-Wahnefried W, Strigo T, Catoe K, Conaway M, Brunetti M, Rimer B, et al. Knowledge, beliefs, and prior screening behavior among blacks and whites reporting for prostate cancer screening. Adult Urology 1995,46:346–51.
(87) Rajaram SS, Rashidi A. Minority women and breast cancer screening: the role of cultural explanatory models. Prev Med 1998;27:757–64.
(88) Lobell M, Bay CR, Rhoads KV, Keske B. Barriers to cancer screening in Mexican-American women. Mayo Clin Proc 1998;73:301–8.
(89) Hubbell FA, Chavez LR, Mishra SI, Valdez RB. Beliefs about sexual behavior and other predictors of Papanicolaou smear screening among Latinas and Anglo women. Arch Intern Med 1996;156:2353–8.
(90) Rimer BK. Putting the "informed" in informed consent about mammography. J Natl Cancer Inst 1995;87:703–4.
(91) Royak-Schaler R, Cheuvront B, Wilson, KR, Williams CM. Addressing women's breast cancer risk and perceptions of control in medical settings. J Clin Psychol Med Settings 1996;3:185–99.
(92) Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995;87:286–92.
(93) Schwartz M, Rimer B, Daly M, Sands C, Lerman C. A randomized trial of breast cancer risk counseling: the impact of self-reported mammography use. Am J Publ Health 1999;89:924–6.
(94) Bowen DJ, Christensen CL, Powers D, Graves DR, Anderson CA. Effects of counseling and ethnic identity on perceived risk and cancer worry in African American women. J Clin Psychol Med Settings 1998;5:365–79.
(95) Lipkus IM, Rimer BK, Biradavolu M. Lowering the perceived risk of breast cancer does not deter women from getting a future mammogram [abstract]. Ann Behav Med 1998;(suppl):S137.
(96) Hiatt RA. Behavioral research contributions and needs in cancer prevention and control: adherence to cancer screening advice. Prev Med 1997;26: S11–8.
(97) Snell JL, Buck EL. Increasing cancer screening: a meta-analysis. Prev Med 1996;25:702–7.

Notes

[1]Although we do not describe how different theories might lead to different hypotheses for predicting screening behavior, it is possible to do so. Weinstein *(23)* outlines circumstances in which it is possible to test among different theories of health-protective behavior, a strategy that we recommend.

[2]We have focused on family history as a risk factor that can prompt increased risk perceptions. In theory, however, any risk factor should be associated with increased screening rates, as long as people are aware of the risk factor/cancer relationship. Cody and Lee *(27)*, for example, found that fairer skin, a fairly well-known predictor of skin cancer risk, predicted whether college students examined their skin for skin cancer.

[3]Screening in relation to symptoms is sometimes diagnostic and is likely different in many ways from routine cancer screening. However, symptoms can prompt screening motives, such as risk and worry, that are similar to those caused by other individual characteristics (e.g., a family history). So, for example, a woman with a prior history of "breast problems" may continue to seek out mammography screening because that prior history increases her perceived risk and associated worry.

Supported in part by Public Health Service grant CA58659 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

We thank Becky Johnson, Tanna Richard, and Katie Quinlan for helping with the literature review.

Risk Communication in Genetic Testing for Cancer Susceptibility

Robert T. Croyle, Caryn Lerman

Risk communication is an integral part of genetic counseling and testing for cancer susceptibility. This paper reviews the emerging literature on this topic. Three relevant aspects of risk communication are addressed: communication of individual risk, communication of the risks inherent in genetic testing, and family communications related to risk. These studies suggest that (*a*) most individuals with some family history of cancer, including those at low to moderate risk, overestimate their personal cancer risk; (*b*) awareness of the risks of genetic testing is limited; (*c*) decision making about genetic testing is influenced strongly by exaggerated perceptions of personal cancer risk and less so by perceptions of the risks of genetic testing; (*d*) perceptions of personal risk of cancer are resistant to standard education and counseling approaches; (*e*) psychologic distress and coping processes influence the processing of risk information and subsequent decision making in genetic testing; and (*f*) family influences play an important role in risk awareness, genetic testing decisions, and outcomes. To study these issues further, new theoretical models and measures of risk perceptions need to be developed. Both observational and experimental methods should be used to examine both the content and process of risk communication in cancer genetic counseling and testing. Emotional, familial, and sociocultural influences on the risk communication process require special attention. [Monogr Natl Cancer Inst 1999;25:59–66]

The isolation of several major cancer susceptibility genes has made it possible to provide predictive genetic testing for cancer susceptibility. Individuals who inherit mutations in these susceptibility genes can be identified years before cancer develops and can engage in appropriate risk-management strategies. Standard protocols for genetic counseling and testing rely heavily on risk communication to convey information about personal and familial cancer risk and to counsel participants about the risks of genetic testing *(1,2)*. For example, pretest counseling sessions focus on the cancer family history and on individual risk status, including the likelihood that the individual has inherited a cancer-predisposing mutation. Another essential element of pretest counseling is a discussion of the risks of genetic testing. Central features of posttest counseling include discussion of the interpretation of the genetic test result, of individual and family risks for developing cancer, and of options for risk management.

Accurate risk comprehension among participants in genetic counseling programs may be critical to their decision making about whether to have a genetic test and, among those who test positive, to their decision making about risk management. Therefore, risk perception and risk communication related to cancer genetics have become active areas of investigation in recent years. Of particular relevance are questions concerning how individuals and families perceive their inherited risks of cancer and how well they understand the risks of genetic testing. Also key are questions concerning the relationship of risk perceptions to genetic testing decisions and the effect of genetic counseling on risk comprehension.

The objective of this paper is to review the emerging data on risk perception and risk communication in genetic counseling and testing for cancer risk. First, we discuss the unique aspects of genetic testing for cancer susceptibility that distinguish this work from other research on risk communication and cancer. Next, we review existing literature on risk perception and communication in cancer genetic counseling and testing. Three relevant aspects of risk communication are addressed: communication of individual risk, communication of the risks inherent in genetic testing, and family communications related to risk. Because this field of research is in the very early stages, we include in our discussion research on risk communication in persons with a family history of cancer, as well as studies of genetic counseling and testing. An important extension of risk communication research, concerning the role of emotion in risk perception and risk communication, is discussed in the following section. Last, we suggest topics for future research to further our understanding in this area.

Unique Aspects of Genetic Testing

Inherited Genetic Risk Is Imposed

Unlike acquired cancer risk factors, such as unhealthy diet, physical inactivity, and smoking, exposure to an inherited mutation cannot be altered by the at-risk individual. In addition, the fact that individuals bear a substantial risk of cancer through no choice of their own may exacerbate a sense of fatalism regarding their future health *(3)*. Experiments designed to examine psychologic reactions to notification of imposed risk factors (as opposed to risk factors resulting from the individual's behavior) have shown a tendency for individuals to minimize the perceived seriousness of the risk *(4)*. Such distortions of risk information may reduce distress in the short term *(5)*, but a sense of fatalism may also diminish motivation to engage in healthy behaviors *(6)*. Individuals may also exaggerate risk-factor information, as documented below in studies of genetic testing for cancer susceptibility *(7)*.

Efficacy of Risk Reduction Strategies Is Unproven

When carriers of cancer-predisposing mutations are advised about options for cancer screening and risk reduction, they also should be informed that these methods have not yet been proven

Affiliations of authors: R. T. Croyle, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD; C. Lerman, Lombardi Cancer Center, Georgetown University Medical Center, Washington, DC.

Correspondence to: Robert T. Croyle, Ph.D., National Institutes of Health, MSC 7326, Executive Plaza North, Rm. 211, Bethesda, MD 20892-7326 (e-mail: bc136e@nih.gov).

See "Note" following "References."

© Oxford University Press

effective for mutation carriers *(8)*. For example, although mammography is effective in older women, its benefits for younger high-risk women and mutation carriers are less certain. The efficacy of screening tests for ovarian cancer is limited for women of all ages, regardless of mutation status *(9,10)*. Even prophylactic surgery, an option chosen by some carriers of cancer predisposing mutations, does not completely eliminate the risk of cancer development *(11)*.

The controllability of a health threat, both perceived and actual, plays an important role in a person's reactions to threatening risk communications *(7,12)*. Individuals who perceive that they can exert control over health outcomes have better adjustment to risk and illness *(13,14)*. In addition, perceived control over illness has been shown to improve medical outcomes by motivating protective health behaviors *(15,16)*. However, when control over health outcomes is uncertain, as is true for inherited risk of cancer, individuals may be less likely to engage in recommended health practices *(7)*. As discussed below, research on risk communication must therefore take into account both actual and perceived control over cancer development to elucidate the optimal approaches to enhance genetic testing outcomes.

Genetic Testing Has Implications for Families

Inherited risk for cancer is shared within families. Typically, genetic testing for cancer risk is initiated with an affected individual in the family. If a known risk-conferring mutation is identified, then other relatives can be tested for this mutation. A common process, employed in most clinical research settings, is to discuss with the individual the implications of his or her test result for other family members *(17)*. This individual is then given the option to contact his or her relatives directly or to have the health care provider contact the relatives. Thus, the individual's communication of his or her test result and risk status is necessary for other family members to have access to testing. Although family members may or may not elect to participate in genetic counseling and testing themselves, the knowledge of their relative's test result may help to clarify their personal risk status. Family communication can also have important effects on psychologic well-being, as has been shown among families coping with other forms of inherited illness *(18)*. As suggested below, family variables are likely to moderate the effect of risk communication on decisions about cancer genetic testing and on quality of life in persons who are tested.

Individual Risk Perception in the Genetic Counseling Context

Association of Perceived Risk to Actual Risk

One of the most consistent findings in behavioral research on cancer risk is that a surprising number of people overestimate their risk for cancer and lack knowledge of risk factors and genetic inheritance. This finding has been documented in studies, ranging from population-based surveys about genetic testing *(19)* to research on hereditary cancer families *(20)*. An earlier study *(21)* examined risk perceptions among 200 first-degree relatives of breast cancer patients who were at low to moderate risk of breast cancer. Their perceived risk of breast cancer, rated on a 1–100 scale, was compared with a lifetime breast cancer probability calculated on the basis of the Gail Model *(22)*. Less than 10% of women reported subjective risk estimates that were within 10 percentage points of their actual risk, and only 2% underestimated their risks. About two thirds of women rated their risk higher than the highest possible risk score that could be achieved for someone in their age group. Overall, the average calculated risk for these women was about 14%, compared with an average perceived risk of breast cancer of about 50%. A more recent study *(23)* classified women from a random-mailed survey into four family history groups (none, minimal, moderate, or strong) on the basis of the model by Claus et al. *(24)*. Over one fourth of women with no family history of breast cancer reported that they had a lifetime risk of breast cancer greater than or equal to 25%. Women in all groups overestimated their lifetime risks. In another study of diverse groups of women with a family history of breast cancer, perceived risks of breast cancer ranged from 3.7-fold to 4.5-fold higher than their actual breast cancer risks *(25)*. Similar findings have been reported for relatives of colon cancer patients *(26)*. In this study population, the average perceived lifetime risk of colon cancer was greater than 50%, despite the fact that the majority of participants had only one affected first-degree relative.

Exaggerated perceptions of personal risk are not limited to lower risk populations. In a study of members of hereditary breast–ovarian cancer families *(20)*, perceived risks of having an altered BRCA1 gene were higher than their actual risk on the basis of the cancer family history. For example, about one third of unaffected females with an affected first-degree relative believed that their risk of inheriting a mutation was 75% or greater, compared with an actual risk of 50%. Although women generally had higher perceived risks than male family members (despite the fact that the probabilities of inheriting a mutation are unaffected by gender), about one half of men without an affected relative reported that their risk was 50% or greater. In a more recent study *(27)*, 60% of women overestimated their risks of being a BRCA1 or BRCA2 mutation carrier. The average difference between perceived risk of being a carrier and estimated risks calculated by use of a risk probability model *(28)* was 32%.

Effect of Cancer Risk Communication on Perceived Risk

How effective are education and counseling in altering exaggerated cancer risk perceptions? To date, this question has been addressed only in the context of breast cancer risk counseling programs, and data on genetic testing for cancer risk are not yet available. In a prospective study, Alexander et al. *(29)* studied 59 women participating in the Breast Cancer Prevention Trial of tamoxifen for chemoprevention of breast cancer. Similar to the previously cited study by Lerman et al. *(21)*, women's average perceived lifetime breast cancer risk was about 50%. However, among 29 women who received individualized risk information (based on the Gail Model) and returned for a follow-up visit, the average perceived risk had dropped to 25%. Although the reduction in risk overestimation was impressive, this figure still greatly exceeded participants' actual risks (15%).

In a retrospective study conducted in the U.K., Lloyd et al. *(30)* compared the risk perceptions of 62 breast cancer risk counseling participants with 62 matched attenders of a general health clinic (controls). Participants in risk counseling perceived themselves to be at greater risk of breast cancer than control subjects; however, this could be the result of a bias in terms of self-selection to risk counseling programs. Despite participation in risk counseling based on the Cancer and Steroid Hormone model, two thirds of women could not provide accurate assessments of their lifetime breast cancer risks. Using a prospective

study design, this research group *(31)* reported on changes in breast cancer risk perceptions following risk counseling among 282 women with a family history of breast cancer. Overall, 9% of women had accurate risk perceptions at baseline, compared with 31% following counseling and 17% at 1-year follow-up. Thus, while there was a small improvement in the accuracy of perceived lifetime risk, the majority of women continued to overestimate their risks. Furthermore, the initial improvement in risk estimation was not maintained 1 year later.

Two studies in the literature used a randomized trial design to evaluate breast cancer risk communication efforts. In one of these studies *(21)*, 200 women were randomly assigned to receive individualized breast cancer risk counseling, based either on the Gail Model or on general health counseling (control group). Although the concordance between subjective and objective breast cancer risk was improved in the risk counseling group, compared with the control group, the magnitude of this effect was small. Similar to the results of Watson et al. *(31)*, about two thirds of subjects in both groups continued to overestimate their risks substantially following both types of counseling. Another study *(32)* tested the provision that an audiotape of the session would improve recall of risk information. One hundred fifteen women with a family history of breast cancer were randomly assigned to receive breast cancer risk counseling either with or without an audiotape. The results showed that provision of the audiotape significantly reduced worries about cancer but did not influence the accuracy of perceived risk.

Perceived Personal Risk and Decision Making About Genetic Testing

Initial studies, conducted before the availability of genetic testing, explored the association of perceived risk with intentions to have a genetic test for cancer. Although these studies used hypothetical scenarios to assess interest in testing and therefore may not be applicable to actual test use, the data on risk perception are of interest. Studies of intentions to have a genetic test for breast cancer *(20,27,33)* have documented universally high interest in testing, with about 70% or more respondents reporting positive intentions to be tested. In these studies, perceived risk, but not actual risk, was found to correlate with genetic testing intentions. Similarly, Lipkus et al. *(34)* found that concern about breast cancer was a more powerful predictor of testing intentions than knowledge of risk factors or actual risk status based on family history.

Similar findings have been reported for colon cancer genetic testing. Interest in genetic testing has been expressed by more than 80% of the respondents to a statewide survey *(35)*, about 90% of colon cancer patients *(36)*, and up to 77% of first-degree relatives *(37–39)*. In a study by Petersen et al. *(39)*, interest in genetic testing was found to be independent of actual risk status based on the number of affected first-degree relatives; however, Glanz et al. *(38)* found that both actual and perceived risk correlated with testing intentions.

These studies have been extended by others that examine the determinants of actual use of genetic counseling and testing. For example, Geller et al. *(40)* found that, among women at high risk for breast cancer, genetic counseling attenders had higher levels of perceived risk. In the setting of genetic testing for colon cancer, perceived risk of cancer, but not actual risk, was found to predict uptake *(26)*. It is also important to note that rates of interest in genetic testing reported in studies using hypothetical scenarios *(20,27,33)* greatly exceed actual test uptake in these studies.

Decision making regarding BRCA1 testing was examined in a study *(41)* of 400 low-to-moderate risk women with a family history of breast or ovarian cancer. The purpose of the study was to test the impact of genetic education and counseling on knowledge, perceived risk, and intention to be tested. Participants were randomly assigned to receive standard pretest education for BRCA1 testing, pretest education plus counseling, or a wait-list control. The education approach reviewed information about personal risk factors; inheritance of cancer risk; and the benefits, limitations, and risks of genetic testing. The education-plus-counseling approach included this information and addressed psychosocial issues, including personal experiences with cancer and the potential psychologic and social effect of testing. At 1-month follow-up, both the education and education-plus-counseling groups exhibited superior knowledge about cancer risk and genetic testing than did the control group. Both counseling groups also showed reductions in perceived risk of having a BRCA1/2 mutation; however, the magnitude of these effects was small. Moreover, neither education approach produced changes in intentions to be tested. Thus, genetic counseling may not greatly diminish perceptions of cancer risk or interest in testing, even among persons at low risk of developing cancer.

Whereas data on individual risk perception and communication are still emerging, the studies described above lead to a few tentative conclusions. First, inflated perceptions of inherited risk of cancer are pervasive in individuals at all levels of risk and for different cancer types. Second, these exaggerated risk perceptions appear to drive testing decisions to a greater degree than does actual risk status. Third, and perhaps most important, perceptions of cancer risk have been refractory to education and counseling.

Communication of the Risks of Genetic Testing

Although perceived risk of cancer is an important component of genetic counseling, information about the benefits and risks of testing may be more essential for informed decision making *(42)*. Despite this finding, there has been relatively little empirical research on this topic. Much of what we know about the perceived risks of genetic testing is derived from studies of attitudes and intentions to be tested rather than from studies of actual participants in genetic testing programs. Studies of potential test participants generally assess the perceived importance of the risks and benefits of genetic testing, based on theories of health behavior *(43)*. For example, Jacobsen et al. *(44)* examined the association of perceived risks and benefits of testing with intentions to have a BRCA1 test among women at risk for familial cancer. The measure of testing benefits included statements relating to the potential positive outcomes of testing (e.g., reassurance, information to facilitate surgery decisions, etc.). A measure of testing risks included statements about the possible negative outcomes (e.g., increases in worries about cancer with a positive test result, insurance discrimination and stigmatization, and negative reactions in family members). As predicted, women who had higher ratings of the risks of genetic testing, relative to their ratings of the benefits, were less likely to intend to be tested. Likewise, Geller et al. *(40)* demonstrated that fear of stigma deterred participation in genetic counseling among women at risk for breast cancer. However, in a study of colon cancer patients who had provided a blood sample *(36)*, the

perceived benefits of testing had a stronger influence on desires to be tested than did perceptions of the risks of testing.

The role of perceptions of genetic testing risks in actual testing decisions was examined in a prospective cohort study of BRCA1/2 testing *(45)*. Overall, the risks of genetic testing were rated as much less important than the benefits. After controlling for demographic and medical factors, uptake of genetic testing was associated with the perceived benefits of testing but not with the perceived risks. This finding underscores the importance of emphasizing communication about the risks of genetic testing to a greater degree in informed consent encounters.

Although these observational studies shed some initial light on the role of perceived risks of genetic testing, experimental designs can provide a stronger test of the effect of risk communications. In the randomized trial of pretest education approaches described above *(41)*, standard education that included the risks of genetic testing had no more effect on participants' perceived risks of genetic testing than the wait-list control condition. However, the education-plus-counseling approach did lead to statistically significant, but modest, increases in perceived risks of testing. This finding suggests that discussion of the psychosocial aspects of genetic testing may have made the risks of testing, especially the psychologic risk, more salient to participants. However, as noted above, neither education nor education plus counseling appeared to influence genetic testing decisions.

Thus, in contrast to the studies showing the overestimation of personal risk of cancer, the results of the above studies show that the risks of genetic testing itself appear to be underestimated. Moreover, perceptions of genetic testing risks exert relatively little influence on genetic testing decisions. Thus, other individual and family factors must exert important influences in the decision-making process.

Risk Communication in Families

One of the most pressing, yet least studied, issues in genetic testing for cancer susceptibility concerns family communication of cancer risks *(46)*. As described above, inherited risk is shared within families, and the risk status of one family member has implications for the others. Thus, familial risk communication is a potentially important determinant of decision making and testing outcomes. This is supported by studies showing that the desire to have genetic testing for colon cancer risk is related to the desire to help family members *(36)* and to levels of family support *(38)*. For high-risk women considering genetic testing for breast cancer risk, having a daughter predicts participation in genetic counseling *(40)*. In this study, participants cited the desire to help relatives as the factor that motivated them to involve other relatives in testing.

Data are beginning to emerge with respect to communication of genetic testing results within families. Green et al. *(47)* found that 88% of breast–ovarian cancer genetic counseling participants shared their consultation summary letter with at least one first-degree relative. In a study of members of hereditary breast cancer families participating in BRCA1/2 testing, 81% of carriers and 85% of noncarriers who had sisters reported that they shared their test results with them *(48)*. Relatively fewer carriers and noncarriers shared results with a brother (60% and 63%, respectively). A higher rate of communicating test results to female than to male relatives was also found by Green et al. *(47)*. This finding may be attributable to the belief that risk information about breast cancer is more relevant to females than to males. Surprisingly, 37% of carriers and 35% of noncarriers reported that they communicated their test results to a child under age 13 years. Although most individuals may disclose their genetic test results to family members, many are reluctant to have clinicians communicate their results to others, at least not without their explicit consent. In a survey of attitudes about BRCA1/2 testing among relatives of breast cancer and ovarian cancer patients, over 80% of respondents reported that health care providers should not disclose their tests results to immediate family members without their written consent *(49)*.

Risk communication within families can have important implications for the psychologic well-being of family members. In a preliminary study of participants in a clinic-based BRCA1/2 testing program *(50)*, BRCA1/2 carriers who communicated their test results to their sisters exhibited a small decrease in psychologic distress, whereas those who did not disclose their results showed a small increase. In the first study to examine the psychological effects of BRCA1 test results among family members, Smith et al. *(51)* found that individual's short-term reactions to testing depended on their siblings' test results. Specifically, among female mutation carriers, psychologic distress was higher among those who were the first in their families to be tested and among those who had siblings who tested negative. For males, psychologic distress was elevated for those who tested negative but who had siblings who all tested positive. The data are consistent with the notion of "survivor guilt" among males, a phenomenon frequently discussed by clinicians but rarely studied empirically. Unfortunately, these studies did not examine the content or process of risk communication in families. Nonetheless, they underscore the point that family communication of risk and its consequences are likely to have an important influence on the decisions and adjustment of individuals and their family members.

Role of Emotion and Coping in Risk Perception and Communication

Both theory and research suggest that emotional factors can modify the cognitive processing of risk-related information when an individual is faced with a personally relevant health threat *(4,52)*. This modification is especially common in risky decisions that are made under emotional stress *(53)*. Given the possibility of receiving very bad news, one would expect the genetic testing situation to be particularly stressful and distressing. According to Janis and Mann *(53)*, stress (or distress) interferes with one's ability to consider the most salient features of the situation and to deliberate carefully about the pros and cons of alternate options. Previous studies provide some support for this conceptualization. In the laboratory setting, anxiety and stress have been found to lead to a narrowing of focus and to impulsive or careless decision making *(54,55)*. In the randomized trial of breast cancer risk counseling described above, the intervention was not successful in improving risk comprehension among participants with higher levels of breast cancer-related distress *(21)*.

In the genetic testing context, psychologic distress appears to have both direct effects on decision making and effects that interact with risk perceptions. In the studies described above, worries about cancer were positively related to interest in genetic testing for breast cancer *(25)* and colon cancer *(36,38,39)*. Cancer-related distress has also been shown to motivate actual use of

genetic testing in high-risk populations. Among members of hereditary breast cancer families, those who reported more frequent intrusive thoughts and worries about cancer were about three times more likely to obtain BRCA1/2 testing *(56)*. Interacting effects of cancer worries and risk perceptions on genetic test use were reported by Codori et al. *(26)*. In this study of individuals at risk for colon cancer, perceived cancer risk predicted use of testing for hereditary nonpolyposis colorectal cancer only in persons who also had high levels of worries about cancer.

The positive association between cancer-related distress and test uptake provides an interesting contrast to findings related to distress and use of predictive testing for Huntington's disease (HD), a progressive neurodegenerative disease. van der Steenstraten et al. *(57)* found that, among individuals at risk for HD, those who were distressed were significantly *less* likely to participate in genetic testing. Specific distress symptoms that predicted nonparticipation included pessimism, preoccupation with HD, and anxiety about adverse outcomes of receiving HD test results. Thus, HD-specific distress and cancer-specific distress appear to exert opposite effects on use of genetic testing. To understand and interpret this discrepancy, one needs to consider key differences between HD and breast–ovarian cancer. Unlike breast–ovarian cancer, a positive genetic test result for HD is associated with a 100% probability of disease. In addition, there are no strategies available to alter the likelihood or course of HD. Thus, disease-specific distress may promote genetic testing only if this action is perceived as offering the potential for risk reduction or altering the course of the disease.

Whereas the studies of cancer-specific distress show a positive relationship with test intentions and uptake, generalized distress may exert different effects. In a study of members of families with hereditary nonpolyposis colorectal cancer *(58)*, depression symptoms were associated with lower rates of test use. In fact, among women, those who reported depression symptoms were about four times less likely to be tested. In comparison with worries about cancer, global distress symptoms may reflect or generate a sense of fatalism about the future that may diminish motivation for health-protective behaviors such as genetic testing.

Thus, the specific effects of distress and emotional factors on genetic testing decisions and outcomes may depend on the level and type of distress and on beliefs regarding the potential for controlling risk. When an individual perceives the outcomes as controllable (i.e., risk can be reduced), cancer-specific distress may motivate active coping strategies, including use of genetic testing. This has been reported for the use of mammography screening in women at risk *(59)*. However, global distress, a perceived lack of control, or both may generate avoidance of genetic testing. This relationship may generate a vicious cycle in which distress generates avoidance that, in turn, generates more distress. This finding is supported by a study of the role of baseline distress in adjustment following BRCA1/2 testing in hereditary breast cancer families *(60)*. In this study, individuals who had high levels of cancer-related distress at baseline but declined to have BRCA1/2 testing reported the highest levels of depression symptoms at 6-month follow-up. Depression symptoms were reported more frequently in these individuals than in relatives who had received positive test results. Post-hoc analyses suggested that these test decliners were motivated to be tested to reduce levels of distress; however, because of fears of discrimination and other adverse outcomes of testing, they avoided testing. Thus, their cancer risk status remained uncertain, and depression levels increased over time. This finding is consistent with previous studies of HD *(61)*.

Implications for Future Research

These initial findings point to the critical challenges faced by educators and counselors who wish to impart more accurate risk information in genetic testing to enhance informed decision making. An important next step in this line of research will be to apply new theoretical models both to explain these findings and to guide the next set of studies on risk communication. In addition, it is time for investigators to focus on the content and process of communication with specific relevance to genetic testing and to extend this work beyond studies of intentions to research on actual testing decisions and outcomes. Although prospective cohort studies of participants in genetic testing would be informative, there also is a need for theoretically guided experimental studies to further our understanding. The discussion below highlights a few specific suggestions for future research in this area.

A critical need for this research domain is the development and validation of better measures of risk perception. A variety of measures are currently in use, including qualitative assessments (e.g., "compared with other women your age, would you say your risk is much lower, lower, about the same . . .?") and quantitative assessments (e.g., "on a scale from 0 to 100, in which 0 equals no chance of getting breast cancer and 100 means you definitely will get it, which number represents the chance that you will get breast cancer someday?"). Qualitative assessments of comparative risk are problematic because the meaning of "lower" or "higher" risk to a given woman is uncertain. Also, because numeric risk figures (or ranges) generally are provided in genetic testing, qualitative assessments may be less sensitive for measuring the effect of counseling on risk comprehension. Whereas an absolute quantitative assessment on a 0–100 scale is appealing because it is consistent with probability information presented by counselors, this assessment is problematic. Two studies that used this measure *(21,29)* both found an average perceived lifetime risk of about 50%. However, anecdotal reports of our study subjects *(21)* suggest that this is interpreted to mean "a 50–50 chance—maybe I'll get it and maybe I won't," thus reflecting uncertainty rather than a specific probability.

Whereas the above measures of risk focus on lifetime cancer risks, it may be more useful to use measures of more immediate relevance in the genetic testing context. For example, an important component of risk in decision making about whether to have a genetic test is the likelihood that a person has an inherited mutation. Such methods have been used in some studies *(20,27)*, although their sensitivity to education and counseling has not yet been tested. In addition, because posttest counseling includes communication of age-specific penetrance (i.e., the likelihood that a mutation carrier will develop cancer by a particular age), risk measures in age increments may be more applicable (e.g., the chances that you will develop breast cancer in the next 5 years). Work described elsewhere in this monograph *(62)* suggests that graphical displays of risk information may also provide a promising alternative to verbal risk communication and assessment.

Experimental study designs would provide an optimal approach to evaluate the merits of alternative methods for risk

communication and for the assessment of risk comprehension. Investigators should consider using analogue or laboratory study designs. These methodologies have been used effectively to manipulate different components of risk-related information (e.g., test validity, time frame for disease onset, disease prevalence, and residual risks) and to evaluate the effects on risk comprehension and health-related decisions *(63)*. A more thorough examination of mental representations of inherited cancers would also advance understanding of the relationships among risk perception, health beliefs, emotional processes, and behavior *(64)*. Experimental designs could be adapted for use in the clinical setting; however, there would be somewhat less flexibility. Whereas it may not be feasible or ethical to manipulate the content of risk communication in genetic counseling, the format could be varied. Observational studies of genetic counseling could also examine the processes and content of risk communication and participant–counselor interactions as well as the effect of these variables on genetic testing decisions and outcomes. Innovative educational approaches, such as CD-ROM technology *(65)*, which allow participants to select the amount and type of information desired, could also be used to determine participants' informational preferences and needs. Such research could highlight factors of importance to risk communication that could subsequently be evaluated in experimental studies.

Another important, but understudied, area of risk communication and genetic testing concerns the influence of contextual factors. Specific contextual influences of importance in risk communication include family factors and sociocultural factors. Some preliminary research on family influences is described above; however, much additional work is needed. Although we have learned that a majority of genetic testing participants communicate their risks to other family members, we know very little about the content, process, and impact of these communications. For example, what factors influence decisions to communicate risk information to some relatives and not to others? What level of detail about test results and personal risk is communicated, and how accurate is this information? How do family communications about cancer risk and genetic testing encourage or discourage other family members from being tested? To address these and other questions about family influences in risk communication and genetic testing, better measures of family communications and interactions are needed *(66)*. In addition, investigators are encouraged to use direct observation and coding methods developed for marital and familial interaction research *(67–69)*.

Finally, much of the research on genetic testing for cancer susceptibility has been conducted with predominantly Caucasian populations. The few available studies of other ethnic groups, mostly African-Americans, highlight differences in risk comprehension. For example, in a series of studies, Hughes et al. *(70)* found that, compared with Caucasian high-risk women, African-American women with a family history of breast cancer were less likely to perceive themselves as being at high risk, had more favorable attitudes about genetic testing for breast cancer and less awareness of the risks *(71)*, had different beliefs about test result disclosure to family members *(49)*, and were more responsive to the effects of pretest counseling interventions *(72)*. Unfortunately, we know very little about the specific cultural beliefs that may underlie these differences. Research in more diverse populations is critical to address these issues.

Summary and Conclusions

As shown from this review, with some exceptions, the initial studies of risk communication in the genetic testing context have been largely descriptive. These studies suggest that (*a*) most individuals with some family history of cancer, including those at low to moderate risk, overestimate their personal cancer risk; (*b*) awareness of the risks of genetic testing is limited; (*c*) decision making about genetic testing is influenced strongly by exaggerated perceptions of personal cancer risk and less so by perceptions of the risks of genetic testing; (*d*) perceptions of personal risk of cancer have not been greatly influenced by standard educational approaches; (*e*) psychologic distress and coping processes influence the processing of risk information and subsequent decision making in genetic testing; and (*f*) family influences play an important role in risk awareness, genetic testing decisions, and outcomes.

These findings also point to directions for future research. To develop a better understanding of risk communication in cancer genetic testing, new theoretical models and measures of risk perceptions need to be developed. Both observational and experimental methods should be used to examine both the content and the process of risk communication. Emotional, familial, and sociocultural influences require special attention. Finally, an improved understanding of risk communication in diverse populations is critical to develop genetic counseling methods that are most effective and most sensitive to the perspectives and backgrounds of individual participants.

References

(1) Botkin JR, Croyle RT, Smith KR, Baty BJ, Lerman C, Goldgar DE, et al. A model protocol for evaluating the behavioral and psychological effects of BRCA1 testing. J Natl Cancer Inst 1996;88:872–82.

(2) Peters JA, Stopfer JE. Role of the genetic counselor in familial cancer. Oncology (Huntingt) 1996;10:159–66, 175; discussion 176, 178.

(3) Marteau TM, Senior V. Illness representations after the human genome project: the perceived role of genes in causing illness. In: Petrie KJ, Weinman JA, editors. Perceptions of health and illness: current research and applications. Amsterdam (The Netherlands): Harwood Academic Publishers; 1997. p. 241–66.

(4) Croyle RT, Sun Y, Hart M. Processing risk factor information: defensive biases in health-related judgments and memory. In: Petrie KJ, Weinman JA, editors. Perceptions of health and illness: current research and applications. Amsterdam (The Netherlands): Harwood Academic Publishers; 1997. p. 267–90.

(5) Timko C, Janoff-Bulman R. Attributions, vulnerability, and psychological adjustment: the case of breast cancer. Health Psychol 1985;4:521–44.

(6) Croyle RT. Appraisal of health threats: cognition, motivation, and social comparison. Cognitive Therapy Res 1992;16:165–82.

(7) Baum A, Friedman AL, Zakowski SG. Stress and genetic testing for disease risk. Health Psychol 1997;16:8–19.

(8) Burke W, Daly M, Garber J, Petersen G, Botkin J, Kahn MJE, et al. Recommendations for follow-up care of individuals with BRCA1 and BRCA2 mutations. JAMA 1997;277:997–1003.

(9) Gargano G, Correale M, Abbate I, Falco G, De Frenza N, Lorusso V, et al. The role of tumour markers in ovarian cancer. Clin Exp Obstet Gynecol 1990;17:23–9.

(10) Hosono MN, Endo K, Sakahara H, Watanabe Y, Saga T, Nakai T, et al. Different antigenic nature in apparently healthy women with high serum CA 125 levels compared with typical patients with ovarian cancer. Cancer 1992;70:2851–6.

(11) Struewing JP, Watson P, Easton DF, Ponder BA, Lynch HT, Tucker MA. Prophylactic oophorectomy in inherited breast/ovarian cancer families. Monogr Natl Cancer Inst 1995;17:33–5.

(12) Lerman C, Glanz K. Stress, coping, and health behavior. In: Glanz K,

Lewis FM, Rimer BK, editors. Health behavior and health education. Theory, research, and practice. 2nd ed. San Francisco (CA): Jossey-Bass; 1997. p. 113–38.

(13) Marks G, Richardson JL, Graham JW, Levine A. Role of health locus of control beliefs and expectations of treatment efficacy in adjustment to cancer. J Pers Soc Psychol 1986;51:443–50.

(14) Taylor SE, Helgeson VS, Reed GM, Skokan LA. Self-generated feelings of control and adjustment to physical illness. J Social Issues 1991;47:91–109.

(15) Thompson SC, Spacapan S. Perceptions of control in vulnerable populations. J Social Issues 1991;47:1–21.

(16) Taylor SE, Kemeny ME, Aspinwall LG, Schneider SG, Rodriguez R, Herbert M. Optimism, coping, psychological distress, and high-risk sexual behavior among men at risk for acquired immunodeficiency syndrome (AIDS). J Pers Soc Psychol 1992;63:460–73.

(17) Biesecker BB, Boehnke M, Calzone K, Markel DS, Garber JE, Collins FS, et al. Genetic counseling for families with inherited susceptibility to breast and ovarian cancer. JAMA 1993;269:1970–4.

(18) Thompson RJ Jr, Gustafson KE, Hamlett KW, Spock A. Stress, coping, and family functioning in the psychological adjustment of mothers of children and adolescents with cystic fibrosis. J Ped Psychol 1992;17:573–85.

(19) Andrykowski MA, Munn RK, Studts JL. Interest in learning of personal genetic risk for cancer: a general population survey. Prev Med 1996;25: 527–36.

(20) Struewing JP, Lerman C, Kase RG, Gambaressi T, Tucker M. Anticipated uptake and impact of genetic testing in hereditary breast and ovarian cancer families. Cancer Epidemiol Biomarkers Prev 1995;4:169–73.

(21) Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995;87:286–92.

(22) Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989; 81:1879–86.

(23) Mouchawar J, Byers T, Cutter G, Dignan M, Michael S. A study of the relationship between family history of breast cancer and knowledge of breast cancer genetic testing prerequisites. Cancer Detect Prev 1999;23: 22–30.

(24) Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. Cancer 1994;73: 643–51.

(25) Durfy SJ, Bowen DJ, McTiernan A, Sporleder J, Burke W. Attitudes and interest in genetic testing for breast and ovarian cancer susceptibility in diverse groups of women in western Washington. Cancer Epidemiol Biomarkers Prev 1999;8:369–75.

(26) Codori AM, Petersen GM, Miglioretti DL, Larkin EK, Bushey MT, Young C, et al. Attitudes toward colon cancer gene testing: factors predicting test uptake. Cancer Epidemiol Biomarkers Prev 1999;8:345–52.

(27) Bluman LG, Rimer BK, Berry DA, Borstelmann N, Iglehart JD, Regan K, et al. Attitudes, knowledge, and risk perceptions of women with breast and/or ovarian cancer considering testing for BRCA1 and BRCA2. J Clin Oncol 1999;17:1040–6.

(28) Berry DA, Parmigiani G, Sanchez J, Schildkraut J, Winer E. Probability of carrying a mutation of breast–ovarian cancer gene BRCA1 based on family history. J Natl Cancer Inst 1997;89:227–38.

(29) Alexander NE, Ross J, Sumner W, Nease RF Jr, Littenberg B. The effect of an educational intervention on the perceived risk of breast cancer. J Gen Intern Med 1995;11:92–7.

(30) Lloyd S, Watson M, Waites B, Meyer L, Eeles R, Ebbs S, et al. Familial breast cancer: a controlled study of risk perception, psychological morbidity and health beliefs in women attending for genetic counseling. Br J Cancer 1996;74:482–7.

(31) Watson M, Lloyd S, Davidson J, Meyer L, Eeles R, Ebbs S, et al. The impact of genetic counseling on risk perception and mental health in women with a family history of breast cancer. Br J Cancer 1999;79:868–74.

(32) Watson M, Duviver V, Walsh MW, Ashley S, Davidson J, Papaikonomou M, et al. Family history of breast cancer: what do women understand and recall about their genetic risk? J Med Genet 1998;35:731–8.

(33) Lerman C, Daly M, Masny M, Balshem A. Attitudes about genetic testing for breast–ovarian cancer susceptibility. J Clin Oncol 1994;12:843–50.

(34) Lipkus IM, Iden D, Terrenoire J, Feaganes JR. Relationship among breast cancer concern, risk perceptions, and interest in genetic testing for breast cancer susceptibility among African-American women with and without a family history of breast cancer. Cancer Epidemiol Biomarkers Prev 1999; 8:533–9.

(35) Croyle RT, Lerman C. Interest in genetic testing for colon cancer susceptibility: cognitive and emotional correlates. Prev Med 1993;22:284–92.

(36) Vernon SW, Gritz ER, Peterson SK, Perz CA, Marani S, Amos CI, et al. Intention to learn results of genetic testing for hereditary colon cancer. Cancer Epidemiol Biomarkers Prev 1999;8:353–60.

(37) Lerman C, Marshall J, Audrain J, Gomez-Caminero A. Genetic testing for colon cancer susceptibility: anticipated reactions of patients and challenges to providers. Int J Cancer (Pred Oncol) 1996;69:58–61.

(38) Glanz K, Grove J, Lerman C, Gotay C, Le Marchand L. Correlates of intentions to obtain genetic counseling and colorectal cancer gene testing among at-risk relatives from three ethnic groups. Cancer Epidemiol Biomarkers Prev 1999;8:329–36.

(39) Petersen GM, Larkin E, Codori AM, Wang CY, Booker SV, Bacon J, et al. Attitudes toward colon cancer gene testing: survey of relatives of colon cancer patients. Cancer Epidemiol Biomarkers Prev 1999;8:337–44.

(40) Geller G, Doksum T, Bernhardt BA, Metz SA. Participation in breast cancer susceptibility testing protocols: influence of recruitment source, altruism, and family involvement on women's decisions. Cancer Epidemiol Biomarkers Prev 1999;8:377–83.

(41) Lerman C, Biesecker B, Benkendorf JL, Kerner J, Gomez-Caminero A, Hughes C, et al. Controlled trial of pretest education approaches to enhance informed decision-making for BRCA1 gene testing. J Natl Cancer Inst 1997;89:148–57.

(42) Sharpe NF. Informed consent and Huntington disease: a model for communication. Am J Med Genet 1994;50:239–46.

(43) Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983;51:390–5.

(44) Jacobsen PB, Valdimarsdottier HB, Brown KL, Offit K. Decision-making about genetic testing among women at familial risk for breast cancer. Psychosom Med 1997;59:459–66.

(45) Lerman C, Narod S, Schulman K, Hughes C, Gomez-Caminero A, Bonney G, et al. BRCA1 testing in families with hereditary breast–ovarian cancer: a prospective study of patient decision-making and outcomes. JAMA 1996; 275:1885–92.

(46) Bowen DJ, Patenaude AF, Vernon SW. Psychosocial issues in cancer genetics: from the laboratory to the public. Cancer Epidemiol Biomarkers Prev 1999;8:326–8.

(47) Green J, Richards M, Murton F, Statham H, Hallowell N. Family communication and genetic counseling: the case of hereditary breast and ovarian cancer. J Genet Counsel 1997;6:45–60.

(48) Hughes C, Lynch H, Durham C, Snyder C, Lemon S, Narod S, et al. Communication of BRCA1/2 test results in hereditary breast cancer families. Cancer Res Ther Control 1999;8:51–9.

(49) Benkendorf JL, Reutenauer JE, Hughes CA, Eads N, Willison J, Powers M, et al. Patients' attitudes about autonomy and confidentiality in genetic testing for breast–ovarian cancer susceptibility. Am J Med Genet 1997;73: 296–303.

(50) Lerman C, Peshkin BN, Hughes C, Isaacs C. Family disclosure in genetic testing for cancer susceptibility: determinants and consequences. J Health Care Law Policy 1998;1:353–72.

(51) Smith KR, West JA, Croyle RT, Botkin JR. Familial context of genetic testing for cancer susceptibility: moderating effect of siblings' test results on psychological distress one to two weeks after BRCA1 mutation testing. Cancer Epidemiol Biomarkers Prev 1999;8:385–92.

(52) Leventhal H, Safer MA, Panagis DM. The impact of communications on the self-regulation of health beliefs, decisions, and behavior. Health Educ Q 1983;10:3–29.

(53) Janis IL, Mann L. Decision making: a psychological analysis of conflict, choice, and commitment. New York (NY): The Free Press; 1977.

(54) Gillis JS. Effects of life stress and dysphoria on complex judgments. Psychol Rep 1993;72:1355–63.

(55) Keinan G. Decision making under stress: scanning of alternatives under controllable and uncontrollable threats. J Pers Soc Psychol 1987;52: 639–44.

(56) Lerman C, Schwartz MD, Lin TH, Hughes C, Narod S, Lynch HT. The

influence of psychological distress on use of genetic testing for cancer risk. J Consult Clin Psychol 1997;65:414–20.

(57) van der Steenstraten IM, Tibben A, Roos RAC, van de Kamp JJP, Niermeijer MF. Predictive testing for Huntington disease: nonparticipants compared with participants in the Dutch program. Am J Hum Genet 1994;55: 618–25.

(58) Lerman C, Hughes C, Trock BJ, Myers RE, Main D, Bonney A, et al. Genetic testing in families with hereditary nonpolyposis colon cancer. JAMA 1999;281:1618–22.

(59) McCaul KD, Branstetter AD, Schroeder DM, Glasgow RE. What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. Health Psychol 1996;15:423–9.

(60) Lerman C, Hughes C, Lemon SJ, Main D, Snyder C, Durham C, et al. What you don't know can hurt you: adverse psychologic effects in members of BRCA1-linked and BRCA2-linked families who decline genetic testing. J Clin Oncol 1998;16:1650–4.

(61) Wiggins S, Whyte P, Huggins M, Adams S, Theilmann J, Bloch M, et al. The psychological consequences of predictive testing for Huntington's disease. Canadian Collaborative Study of Predictive Testing. N Engl J Med 1992;327:1401–5.

(62) Lipkus IM, Hollands JG. The visual communication of risk. Monogr Natl Cancer Inst 1999;25:149–63.

(63) Ditto PH, Croyle RT. Understanding the impact of risk factor test results: insights from a basic research program. In: Croyle RT, editor. Psychosocial effects of screening for disease prevention and detection. New York (NY): Oxford University Press; 1995. p. 144–81.

(64) Leventhal H, Benyami Y, Brownlee S, Diefenbach M, Leventhal EA, Patrick-Miller L, et al. Illness representations: theoretical foundations. In: Petrie KJ, Weinman JA, editors. Perceptions of health and illness: current research and applications. Amsterdam (The Netherlands): Harwood Academic Publishers; 1997. p. 19–45.

(65) Green MJ, Frost N. An interactive computer program for educating and counseling patients about genetic susceptibility to breast cancer. J Cancer Educ 1997;12:204–8.

(66) Daly M, Farmer J, Harrop-Stein C, Montgomery S, Itzen M, Costalas JW, et al. Exploring family relationships in cancer risk counseling using the genogram. Cancer Epidemiol Biomarkers Prev 1999;8:393–8.

(67) Cutrona CE. Social support as a determinant of marital quality: the interplay of negative and supportive behaviors. In: Pierce GR, Sarason BR, Sarason IG, editors. Handbook of social support and the family. New York (NY): Plenum; 1996. p. 173–94.

(68) Gottman JM. The roles of conflict engagement, escalation, and avoidance in marital interaction: a longitudinal view of five types of couples. J Consult Clin Psychol 1993;61:6–15.

(69) Pasch LA, Bradbury TN, Sullivan KT. Social support in marriage: an analysis of intraindividual and interpersonal components. In: Pierce GP, Lakey B, Sarason IG, Sarason BR, editors. Sourcebook of social support and personality. New York (NY): Plenum; 1997. p. 229–56.

(70) Hughes C, Lerman C, Lustbader E. Ethnic differences in risk perception among women at increased risk for breast cancer. Breast Cancer Res Treat 1996;40:25–35.

(71) Hughes C, Gomez-Caminero A, Benkendorf J, Kerner J, Isaacs C, Barter J, et al. Ethnic differences in knowledge and attitudes about BRCA1 testing in women at increased risk. Patient Educ Counsel 1997;32:51–62.

(72) Lerman C, Hughes C, Benkendorf JL, Biesecker B, Kerner J, Willison J, et al. Racial differences in testing motivation and psychological distress following pretest education for BRCA1 gene testing. Cancer Epidemiol Biomarkers Prev 1999;8:361–7.

Note

Drs. Croyle and Lerman are joint first authors of this manuscript. The order of their names is entirely arbitrary.

Decision Aids for Patients Considering Options Affecting Cancer Outcomes: Evidence of Efficacy and Policy Implications

Annette M. O'Connor, Valerie Fiset, Catherine DeGrasse, Ian D. Graham, William Evans, Dawn Stacey, Andreas Laupacis, Peter Tugwell

Some cancer screening and treatment decisions are not clear cut because outcomes are uncertain or options have different benefit/risk profiles. "Decision aids" have been developed as adjuncts to counseling so that patients can learn about benefits and risks, can consider their personal values, and can participate with their practitioner in decision making. The purpose of this paper is to review published evidence about the efficacy of decision aids focused on cancer outcomes and to outline research and dissemination issues. Studies evaluating cancer-related decision aids demonstrate that they are acceptable to patients and help those who are uncertain at baseline to make choices. They also increase the likelihood that choices are based on better knowledge, realistic expectations of outcomes, and personal values. Decision aids reduce some dimensions of decisional conflict, and their effect on decisions is variable. Few studies examine the downstream effects of decision aids on long-term persistence with choices, regret, and quality of life. The differences between simpler and more intensive methods of decision support appear to be negligible in terms of knowledge and satisfaction as well as variable in terms of decisions and decisional conflict. However, more intensive methods are superior in terms of user acceptability and of the extent to which choices are based on realistic expectations and personal values. The clinical importance of these differences and the cost-effectiveness remain to be established. On the basis of this review, several recommendations for research are made, and dissemination issues are identified. [Monogr Natl Cancer Inst 1999;25:67–80]

Some cancer screening and treatment decisions are not clear cut because of the uncertainty of the evidence on outcomes or the lack of consensus that the benefits outweigh the risks *(1)*. Practice guidelines for these difficult decisions often recommend that practitioners exercise judgement in applying them to individual patients and that patients' values for the outcomes be considered *(2,3)*. Accordingly, decision support interventions, known as "decision aids" or "shared decision-making programs," are being developed as adjuncts to practitioners' counseling *(4–10)* so that patients can 1) understand the probable benefits and risks of options, 2) consider the value they place on benefits versus risks, and 3) participate actively with their practitioners in deciding about options *(2)*. The purpose of this paper is to review published evidence about the efficacy of decision aids focused on cancer outcomes and to outline research and policy implications. We will define decision aids, examine the reasons for their development, present their theoretical underpinnings, review the evidence of their efficacy, and discuss implications.

What Is a Patient Decision Aid?

Decision aids are used as adjuncts to practitioners' counseling to prepare patients for decision making. According to the Cochrane definition *(11)*, they are interventions designed to help people make specific and deliberative choices among options by providing (at the minimum) information on the options and outcomes relevant to the person's health status. Additional strategies may include information on the disease or condition, probabilities of outcomes tailored to a person's health risk factors, an explicit values-clarification exercise, information on others' opinions, and guidance or coaching in the steps of decision making and communicating with others. Decision aids may be administered with the use of various media, such as decision boards, interactive videodiscs, personal computers, audiotapes, audio-guided workbooks, pamphlets, and group presentations. Excluded from the definition of decision aids are passive informed consent materials, educational interventions that are not geared to a specific decision, or interventions designed to promote compliance with a recommended option rather than a choice based on personal values.

Why Are Patient Decision Aids Developed?

The development of decision aids in several centers in the United States, the United Kingdom, and Canada is motivated by several trends: (*a*) the rise of consumerism with an emphasis on informed choice rather than informed consent, (*b*) the evidence-based practice movement disseminating evidence to consumers as well as to practitioners, (*c*) the interest in consumer-focused strategies to reduce large practice variations caused by supplier-induced demand, (*d*) the identification of treatment decisions that are "utility" or "value" sensitive from decision analyses, (*e*) the proliferation of overviews and outcome studies that provide estimates of outcomes for use in decision aids, and (*f*) the evo-

Affiliations of authors: A. M. O'Connor, School of Nursing and Faculty of Medicine, University of Ottawa, and Clinical Epidemiology Unit, Loeb Health Research Institute, Ottawa Hospital, Ontario, Canada; V. Fiset, School of Nursing, University of Ottawa, Ontario; C. DeGrasse, School of Nursing, University of Ottawa, and Ottawa Hospital Breast Diagnostic Centre, Ontario; I. D. Graham, A. Laupacis, P. Tugwell, Faculty of Medicine, University of Ottawa, and Clinical Epidemiology Unit, Loeb Health Research Institute, Ottawa Hospital, Ontario; W. Evans, Faculty of Medicine, University of Ottawa, and Cancer Care Ontario, Ottawa Regional Cancer Center; D. Stacey, School of Nursing, University of Ottawa, and Cancer Care Ontario, Ottawa Regional Cancer Center.

Correspondence to: Annette M. O'Connor, R.N., Ph.D., Clinical Epidemiology Unit, C-4 Rm. 402, Loeb Health Research Institute at the Ottawa Hospital, Civic Campus, 1053 Carling Ave., Ottawa, ON, Canada K1Y 4E9 (e-mail: aoconnor@LRI.CA).

See "Note" following "References."

© Oxford University Press

lution of patient preference-oriented health policy that reserves interventions for those patients who consider the treatment benefits to outweigh the risks (e.g., reserving palliative surgery for patients who consider symptom relief worth the surgical risks, rather than basing a surgical policy on the average patient's utilities).

Kassirer *(3)* lists some indications for explicitly eliciting patients' preferences in clinical practice: 1) Options have major differences in outcomes or complications, 2) decisions require making trade-offs between near-term and long-term outcomes, 3) one choice can result in a small chance of a grave outcome, or 4) marginal differences are found in outcomes between options. Patient characteristics may also determine the need for a decision aid, e.g., if patients are very risk averse or attach unusual importance to certain possible outcomes.

Another useful strategy for determining the need for a decision aid is to classify treatment policies as standards, guidelines, or options using definitions by Eddy *(1)*. Table 1 summarizes the approach. For standards of care in which outcomes are known and patients' preferences are generally consistent in favoring an intervention, decision aids may be less useful, and conventional informed consent procedures are more appropriate. In contrast, decision aids may be indicated for treatment guidelines or options because outcomes may be more uncertain, or values for the benefits relative to the risks are more variable or unknown. For example, the guidelines *(2)* for postmenopausal hormone replacement therapy (HRT) recommend that decisions be tailored to a woman's hysterectomy status and risk of coronary heart disease (CHD), osteoporosis, and breast cancer. Moreover, a woman needs to consider her values in balancing the potential benefits of reducing the risk of osteoporosis and CHD and relief of menopausal symptoms against the potential risk of endometrial and breast cancers, the side effects of HRT, and her attitudes toward taking medication for a natural aging process.

A recent example of an "option" emerged in 1998 following the early completion of the U.S. National Cancer Institute (NCI) Tamoxifen Prevention Trial. Although two smaller trials demonstrated no benefit, the NCI trial showed that tamoxifen reduced the incidence of breast cancer in high-risk women but also increased the risk of endometrial cancer, pulmonary embolus, and deep-vein thrombosis. Within 6 months of stopping the NCI trial, (*a*) NCI published the results on the Internet, including a statement that the decision about treatment depends on a woman's personal health history and how she weighs the benefits and risks; (*b*) the results of the three prevention trials were published; (*c*) tamoxifen was approved by the Food and Drug Administration (FDA) for use in women at high-risk for breast cancer; and (*d*) a "risk disk" to help practitioners and women identify personal risks for breast cancer and understand benefits and risks of tamoxifen was developed and disseminated. Note that the public and practitioners had simultaneous access to the information.

How Are Decision Aids Designed to Work?

Although the developers of decision aids have different conceptual frameworks of decision support *(12–19)*, most are based on decision theories from economics and cognitive psychology *(20–22)* that structure decisions according to options, outcomes, and probabilities of outcomes so that patients are better able to judge the value of the benefits versus the risks. Many frameworks broaden this cognitive perspective by including emotional, social, or environmental dimensions *(23–27)*. For example, the Ottawa Framework *(12)* identifies several determinants of health care decisions that may be suboptimal and are potentially modifiable by decision aids. Patients and practitioners may have problems with (*a*) perceptions of the decision (e.g., inadequate knowledge, unrealistic expectations of outcomes, unclear values, high uncertainty. or decisional conflict); (*b*) perceptions of others (e.g., biased or limited perceptions of the variation in others' opinions and practices, social pressures, or inadequate support); and (*c*) personal and external resources to make the decision (e.g., limited skills in shared decision making). Decision aids are designed to address these problematic determinants of choice by providing accurate, balanced, and tailored information; by clarifying patients' values; and by augmenting skills in shared decision making.

For example, knowledge may be improved by providing information on options and outcomes. Unrealistic expectations (perceived probabilities of outcomes) may be realigned by presenting probabilities of outcomes that are tailored to the patient's clinical risk and by describing outcomes so that they are easy to imagine and to identify with *(21)*. Unclear values are addressed by describing outcomes in familiar, simple, and experiential terms so as to better judge their value *(22)* and by providing the opportunity to weigh the benefits versus the risks. Biased perceptions of the variation in others' opinions may be corrected by presenting all options and, in some cases, by providing examples of others' choices and statistics on variation in choices. Shared decision-making skills may be improved by providing structure and guidance in deliberating about the personal issues involved in the choice and in communicating preferences *(28–35)*. As a consequence of these interventions, patients presenting with uncertainty or decisional conflict caused by these problems may become more certain about what to choose and may be more likely to implement these choices *(36)*.

On the basis of the Ottawa framework, one can hypothesize that decision aids will improve the determinants of choice so that decisions are more likely to be 1) informed (i.e., based on better

Table 1. Determining need for patient decision aids on the basis of the Eddy (*1*) classification of health policy decisions

	Practice standards	Practice guideline	Practice options
Likelihood of outcomes	Known	Known	Known/unknown
Agreement in patients' values/preferences for outcomes	Known, unanimous (≥95% agreement)	Known, majority (≥60% agreement)	Known, evenly split Unknown
Recommendation for treatment	Yes	Variable	No
Patient participation in decision making	Passive, informed consent	Variable	Active, informed choice
Practitioner intervention needed	Counseling followed by verbal or written consent	When recommendation states the decision should be based on patients preference, patient decision aid can be used with follow-up counseling	Patient decision aid and follow-up counseling

knowledge and realistic expectations), 2) consistent with personal values, and 3) implemented. Moreover, patients' comfort with the decision-making process (e.g., decisional conflict, self-confidence, and satisfaction with decision making) may be improved. On the basis of the results of other educational interventions *(28,29,37–39)* designed to promote realistic expectations of outcomes and informed active involvement in one's care, it is also reasonable to hypothesize that patients may be more likely to persist with decisions, to report less distress with the consequences of their decisions, and to experience improved health-related quality of life.

Do Decision Aids Work?

To date, there have been two published documents describing the efficacy of decision aids: an annotated bibliography *(40,41)* and a report to the Agency for Health Care Policy and Research *(15)*. A Cochrane collaboration systematic overview *(11)* of randomized trials of decision aids is currently in progress. This paper summarizes the evaluative studies from the annotated bibliography *(41)*, including an update to early 1998. The following databases were searched: MEDLINE®, CINAHL®, PsycINFO®, and Current Contents®. We also hand searched *Medical Decision Making* and *Health Expectations*. In this paper, we limited our review to studies of interventions that would be classified as decision support according to the Cochrane definition (*see* previous section). We included three types of studies 1) before/after studies that evaluated decision aids with patients at the point of decision making, 2) randomized trials that evaluated decision aids in comparison to "usual care" with patients at the point of decision making, and 3) randomized experiments comparing different methods of decision support in decision aids either with patients at the point of decision making or with volunteers making hypothetical choices. We included studies of decision aids that were not cancer related because of the small numbers of studies available in this emerging field of research.

Types of Decision Aids Studied

Table 2 summarizes the decisions aids that were identified in this review. Over half of the studies focused on cancer-related topics. Most decision aids focused on surgical or medical therapies, although a few considered preventive, end-of-life, and clinical trial participation decisions. Various media were used for delivering the decision aids.

In some of the studies, it was difficult to ascertain what exactly was done in the "black box" of the decision aid. Although all decision aids included information on the options, benefits, and risks, they varied considerably in their presentation of this information. Moreover, there was considerable variability in whether other decision-support strategies were included in the decision aid. It was also sometimes difficult to ascertain what was involved in "usual care" or in the comparative intervention.

Evaluative Studies of Efficacy

Table 3 summarizes the results of the evaluative studies with patients at the point of decision making in which before/after designs and randomized trials with usual care controls were used.

Effect on Variation in Choices

One of the main rationales for using decision aids has been to reduce "inappropriate" practice variation. It is assumed that, if the asymmetry of information available to practitioners and patients regarding options, outcomes, and patients' values is corrected by decision aids, then choices may more appropriately reflect patients' preferences. The direction of the shift in choices will depend on the cause of the practice variation. For example, if overuse is caused by supplier-induced demand involving knowledgeable practitioners who assume patients' values are similar to their own and uninformed patients, rates of use may decline if patients' informed valuing of options does not correspond to those of their practitioners. If there is underuse of interventions because of uninformed practitioners or patients, rates of use may increase.

The before/after studies *(12,42–45,60)*, including two cancer-related studies *(12,60)*, are consistent in showing that decision aids have the greatest effect on the choices of those who are undecided at baseline. Approximately 18%–30% of patients were undecided before using decision aids, and 44%–68% of the undecided made a choice after using aids. The fact that over a third of the undecided still could not make up their mind after using a decision aid underscores the difficulty of these decisions and the need for follow-up counseling by practitioners.

In contrast, decision aids are less likely to change the decisions of the 70%–82% of individuals who have a stated preference at baseline *(12,42–44)*. Changes occurred in only 5%–13% of those preferring a more intensive treatment at baseline and in 11%–18% of those preferring a less intensive treatment. However, in some studies *(45,54,60)*, the decision aid did shift preferences toward the less intensive option, e.g., toward breast-conserving surgery for breast cancer *(60)*.

The randomized trials comparing decision aids with usual

Table 2. Types of decision aids evaluated

Type of decision (reference No.)*	Medium of delivery*
Medical or surgical treatments	
Cancer, breast (*6,52,60,62,71,73,93*)	**Interactive videodisc, decision board, brochure, interactive mutimedia program, audio-guided workbook**
Cancer, lung (*55,58*)	**Audio-guided workbook, personal interview with trade-off exercise**
Cancer, leukemia (*63*)	**Decision board**
Cancer, lymphoma (*53*)	**Poster**
Atrial fibrillation (*46,56*)	Audio-guided workbook
Benign prostatic hypertrophy (*4,43,45,59*)	Interactive videodisc
Low back pain (*42*)	Interactive videodisc
Circumcision (*50*)	Written materials
Ischemic heart disease (*44,48,49*)	Interactive videodisc, video
Screening or diagnostic tests	
Colon cancer screening (*64,76*)	**Computer program**
Prostate-specific antigen tests (*47,51,69,72*)	**Video, pamphlet, scripted counseling**
Amniocentesis (*9,54*)	Audio-guided workbook
Preventive therapies	
Hormone therapy (*7,8,12,67,68*)	**Audio-guided workbook, brochure, group workshops**
Hepatitis B vaccine (*5*)	Written materials
Clinical trial participation	
Breast cancer therapy (*65*)	**Written vignettes**
Cancer chemotherapy (*70*)	**Audiotape, computer, interactive computer**
End-of-life	
Resuscitation in seniors (*57*)	Written outcome data

*Decision aids focused on cancer outcomes are in bold.

Table 3. Evidence of effectiveness of patient decision aids

	Before/after studies detecting change from baseline after decision aid (reference No.)*	Randomized trials of decision aids vs. usual care controls (reference No.)*
	Do decision aids affect decisions?	
	Yes Help undecided *(**12,60**,42–45)* Reduce preference for intensive option *(**60**,45,54)*	**Yes** Help undecided *(46)* Reduce preference for intensive option *(**47,51**,46)* **Trend** decision aid reduces preference for intensive option *(4,48,49)* **No** *(50)*
	Do decision aids improve determinants of decisions?	
Choice based on better knowledge of options, benefits, or risks	**Yes** *(**8,12,52**,42,43)* and **no** *(57)*	**Yes** *(**47,51**,4,46,48,49)* and **no** *(50)*
Choice based on more realistic expectations (agreement between perceived probability of outcomes and estimates derived from evidence)	**Yes** *(**12,53,55,58**,54,56,57)*	**Yes** *(**47,51**,46)*
Choice based on personal values (correlation between personal values for outcomes and choice)	**Yes** *(**12,58**,59)*	—
	Do decision aids improve comfort with decision making?	
Reduce decisional conflict (feeling uncertain, uninformed, unclear about values, unsupported in decision making)	**Yes** *(**12,55,58,60**,54)*	**Yes**, benefit limited to informed dimension *(46,48)*
More satisfied with decision-making process	—	**Yes** *(4)* and **no** *(46,48,49)*
More satisfied with decision	**Yes** *(58)*	**No** *(4,48,49)*
More satisfied with treatment	—	**Less satisfied** *(49)*
	Do decision aids improve outcomes of decisions?	
Better health-related quality of life	—	**Yes** *(4)* and **no** *(49)*
Better compliance/persistence with decision	—	—
Reduced morbidity	—	—
	Do decision aids produce favorable reactions from patients?	
Acceptable to patients (comprehensibility, length, clarity, usefulness, interest, balance, or recommend to others)	**Yes** *(**6,58,59,60,62**,42–44,57)* **Variable** *(42)*	
Feasible to use	**Yes** *(**63,64**)*	—

*Decision aids focused on cancer outcomes are in bold.

care also suggest that decision aids affect decisions. They reduced the proportion of undecided in one trial *(46)* and showed a trend toward reducing preferences for more intensive options by 22%–48% in six of seven trials.

Flood et al. *(47)* underscore the importance of predisposition and financial incentives in changing the decisions of patients. Two trials *(47,51)* showed that men who were exposed to decision aid about the prostate-specific antigen (PSA) before a routine scheduled fee-for-service visit had close to half the rates of PSA testing than usual-care control subjects. In contrast, the rates of PSA testing in another study *(47)* were comparable in those exposed to a decision aid (98%) versus a general education video (100%) in a clinic where men attended specifically for a free PSA test. However, the men receiving the free test differed in intentions toward having future PSA tests (which they would presumably have to pay for); in the decision-aid group, 74% had strong intentions to have a PSA test compared with 90% in the education video group.

Effect on Determinants of and Comfort With Decision Making

The most important question is: Are these observed changes in decisions after using a decision aid accompanied by commensurate improvements in the determinants of the choices? In Table 3, the before/after studies of decision aids focused on cancer and other outcomes are consistent in demonstrating that patients' choices are more likely to be based on better knowledge *(8,12,42,43,52)*, more realistic expectations of outcomes *(12,53–58)*, and personal values for outcomes *(12,58,59)* after patients use decision aids. Moreover, patients' comfort with their decisions, as measured by the Decisional Conflict Scale, is improved *(12,54,55,58,60)*.

The randomized trials, including two trials of PSA decision aids *(47,51)*, have confirmed that decision aids are superior to usual care in improving knowledge *(4,46–49,51)* and in creating realistic expectations *(46,47,51)*. Decisional conflict is also improved *(46,48)*, but the benefit is confined to feeling more informed about options, benefits, and risks. These results point to one of the main mechanisms explaining the potentially conservative effect of decision aids on decision making. The decision aids are better than usual care at moderating patients' exaggerated perceptions of risk of disease without the intervention and their exaggerated perceptions of the benefits of interventions. The aid also gives them a better appreciation of the potential risks associated with the more intensive intervention. Therefore, fewer patients are likely to judge that the potential benefits of the intensive option outweigh the potential harms.

There have been no trials examining the effect of cancer-related decision aids on patients' satisfaction with decision making. For other trials of decision aids, most have shown no impact.

Effect on Outcomes of Decisions

Although outcomes of decisions are difficult to judge when decisions are based on personal values, it is still useful to ex-

amine how decision aids affect long-term persistence with choices, distress, regret, and health-related quality of life. There are no cancer studies examining the downstream effects of decision aids. The two noncancer studies *(4,49)* examining quality of life had variable results.

Patients' Reactions to Decision Aids

Before/after studies *(6,42–44,57–60,62–64)* are generally consistent in demonstrating that decision aids are acceptable to patients and are feasible to use. Further evaluations are needed to establish acceptability to practitioners and to patient groups who vary by age, education, ethnicity, and preferences for participation in decision making.

Randomized Experiments Comparing Simpler With More Intensive Methods of Decision Support

Table 4 summarizes the trials comparing simpler with more intensive methods of decision support. All but two of the trials focused on decision aids with cancer outcomes. In all of the decision-support interventions being compared, patients were provided with some information on the options and the outcomes, but there may have been differences in the complexity of the medium of delivery; the amount of information on options and outcomes; the use of probabilities; the inclusion of examples of other patients' decision making; and the guidance in deliberation regarding the personal issues, such as perceptions of the probabilities of outcomes, personal values, questions, and choices or leanings. There are too few studies to draw substantive conclusions; therefore, the following summary is considered preliminary.

Although the decision aids are known to improve baseline knowledge *(8,12,72)*, when different methods of decision support were compared with one another, no one method was superior at increasing knowledge *(8,7,52,70,71,73)*. This null result is likely because of the overlap in information provided by the alternative interventions. Possibly, more sensitive knowledge tests may be able to detect differences, but one needs to be careful that the knowledge that is tested is considered essential to patients for decision making.

The methods used in decision aids do affect patients' expectations of outcomes. Patients have more realistic expectations if they are exposed to quantitative presentations of probabilities *(7,65)* and if outcomes are framed *(66)* positively (e.g., chance of remaining free of treatment side effects = 95%) rather than negatively (e.g., chance of having a side effect = 5%). However, the effect on decisions of creating more realistic expectations is more variable. It did dampen patients' enthusiasm for participating in a hypothetical clinical trial *(65)* but had no effect on whether a patient took HRT *(7)* or accepted an influenza vaccine *(66)*. However, those exposed to positive frames reported fewer vaccine side effects and less absenteeism from work *(66)*.

Methods in decision aids do affect whether choices reflect personal values. The correlation between personal values and choices improves when decision aids provide detailed and probabilistic information on outcomes and when patients are asked to deliberate about the personal probabilities and values for each outcome *(5,7,67,68)*. The incremental benefit of asking patients to consider their values is an understudied area; one study *(68)* that used a "weigh scale" values clarification exercise found little overall incremental benefit, except possibly in those patients who were considering a change from the status quo. The effect of values clarification exercises needs to be studied in groups that are actively considering change to confirm these results and to determine if they improve the quality of discourse with practitioners regarding values, as well as if they have an effect on long-term persistence with decisions.

Methods in decision aids do affect some decisions, although it is difficult to determine which components in the decision aid produced these effects; e.g., it is not clear whether decisions are influenced by providing examples of how others make decisions. The two studies *(68,73)* that varied the inclusion of examples and showed a possible effect on choices were confounded by varying other aspects of the decision aid. Moreover, in both of these studies, the decision was hypothetical, thereby limiting the generalizabilty of results to patients who actually face the decision and who use more than the information in an aid to make their decisions. The use of examples is controversial. Some decision aids do not include them to remain "neutral"; others use them to convey the variability in other patients' choices.

Few studies have examined the separate effects of guidance or coaching in deliberation about options and communication of preferences. Coaching has had a beneficial effect on patient outcomes *(28–35)*, and whether it can augment the beneficial effects of decision aid remains to be seen.

Few studies have been conducted that compared only the medium of delivery, usually because some methods of decision support require more complicated delivery technologies. In the two studies that did compare media, one *(70)* showed that an interactive computer program increased entry decisions for a hypothetical clinical trial compared with an audiotape. Another study *(73)* comparing multimedia with a pamphlet found a trend toward decreased preferences for mastectomy but no differences in involvement in decision making.

The different methods used in decision aids had no affect on satisfaction with the decision. There is a need to reexamine the way satisfaction is measured, given its lackluster performance, in discriminating not only between decision aids but also between decision aids and usual care. It may be more appropriate to measure satisfaction with preparation for decision making (which presumably decision aids do well) than satisfaction with the process of decision making and with the practitioner (which depends on many factors outside the control of decision aids). Moreover, investigators may need to acknowledge the difficulty in demonstrating improvements in satisfaction with the decision when choices are inherently difficult to make because of competing benefits and risks. Furthermore, once the decision is made, patients may find it more psychologically comforting to say that they are satisfied with it rather than entertain doubts about what they chose *(61)*. Perhaps a better indication of satisfaction with the decision is persistence with the choice; unfortunately, this indicator is only useful for revocable decisions. In a current trial (O'Connor A et al.: unpublished data), we have had success using a scale that elicits patients' and practitioners' satisfaction with patients' preparation for decision making. The scale discriminates well (effect size 1.8) between a pamphlet decision aid regarding HRT and one with the full range of interventions delivered via audio-guided workbook.

The acceptability of decision aids was affected by methods when the differences between interventions were large. For example, users found that pamphlets (with less detail, no illustrations, no probabilities, no examples, and no guidance in personal

Table 4. Randomized trials comparing different methods in decision aids

Decision (reference No.)*	Delivery*†	Options*†	Outcomes*†	Probabilities*†	Examples*†	Guidance in personal deliberation*†
Hormones *(8,67)*	(*a*) **Pamphlet**	Detailed	Detailed	Yes	**Yes**	**No**
	(*b*) **Group**	Detailed	Detailed	Yes	**Yes‡**	**No**
	(*c*) **Group**	Detailed	Detailed	Yes	**Yes‡**	**Yes—probabilities, values, choice**
Hormones *(7)*	(*a*) **Pamphlet**	**Brief**	**Brief**	**No**	**No**	**No**
	(*b*) **Audio-guided workbook**	**Detailed**	**Detailed**	**Yes**	**Yes**	**Yes—probabilities, values, choice**
Hormones *(68)*	(*a*) Audio-guided workbook	Detailed	Detailed	Yes	**No**	Probabilities, questions
	(*b*) Audio-guided workbook	Detailed	Detailed	Yes	**Yes**	Probabilities, questions, **values, choice**
Breast cancer surgery *(71)*	(*a*) **Pamphlet**	**Brief**	**Brief**	**No**	**No**	**No**
	(*b*) **Audio-guided workbook**	**Detailed**	**Detailed graphic**	**Yes**	**Yes**	**Yes—values, choice**
Hepatitis B Vaccine *(5)*	(*a*) Written	**Brief**	**Brief**	**No**	No	**No**
	(*b*) Written	**Detailed**	**Detailed**	**Yes**	No	**No**
	(c) Written	**Detailed**	**Detailed**	**Yes**	No	**Yes—probabilities, values**
Trial entry *(65)*	(*a*) Vignette	Brief	Brief	**Qualitative**	—	—
	(*b*) Vignette	Brief	Brief	**Quantitative**	—	—
Flu vaccine *(66)*	(*a*) Chart/poster	Detailed	Detailed	**Positive frame**	—	—
	(*b*) Chart/poster	Detailed	Detailed	**Negative frame**		
Transplant *(53)*	(*a*) Chart/poster	Detailed	Detailed	**Numbers**	—	—
	(*b*) Chart/poster	Detailed	Detailed	**Numbers/graph**		
PSA test *(72)*	(*a*) Pamphlet	Brief	Brief	**No information**		
	(*b*) Pamphlet	Brief	Brief	**False positive/negative**		
PSA tests *(69)*	(*a*) Verbal	Sentence	**Very brief**	**None**	No	
	(*b*) Verbal	Sentence	**Detailed**	**Quantitative**	No	
Breast cancer surgery *(52)*	(*a*) **Booklet**	Detailed	Detailed	Yes	**No**	
	(*b*) **Video**	Detailed	Detailed	Yes	**Yes**	
Breast cancer surgery *(73)*	(*a*) **Pamphlet**	**Briefer**	**Briefer**		**No**	
	(*b*) **Multimedia**	**Detailed**	**Detailed**		**Yes**	
Trial *(70)*	(*a*) **Audiotape**	Detailed	Detailed	Yes	No	**No**
	(*b*) **Computer**	Detailed	Detailed	Yes	No	**Choice of order of information**

(Table continues)

*Decision aids focused on cancer outcomes are in bold. HRT = hormone replacement therapy; PSA = prostate-specific antigen.
†Components are bolded to signify differences between methods used.
‡Patients in the two groups were exposed to more examples than those using the brochure.

deliberation) were less acceptable than audio-guided workbooks that included these strategies. However, there was no overall difference in acceptability when more subtle differences in decision aids were compared (e.g., the addition of graphical displays to accompany numerical probabilities or the addition of a weigh scale values clarification exercise).

Conclusions About Evaluative and Methodologic Studies

In the evaluative studies of cancer-related decision aids conducted to date, the aids are acceptable to patients and help those who are uncertain at baseline to make a choice. They also increase the likelihood that choices are based on better knowledge, realistic expectations of outcomes, and personal values. They reduce some dimensions of decisional conflict, and their effect on decisions is variable. Few studies examine the downstream effects of decision aids on long-term persistence with choices, regret, and quality of life.

In terms of methods used in decision aids, there has been minimal investigation of what works in the "black box" of decision aids. When simpler methods are compared with more intensive methods of decision support, the differences are negligible in terms of knowledge and satisfaction and are variable in terms of decisions and decisional conflict. However, more intensive methods are generally superior in terms of user acceptability and the extent to which choices are based on realistic expectations and personal values. The clinical importance of

Table 4 (continued). Randomized trials comparing different methods in decision aids

Choice	Knowledge	Expectations	Correlation values and choice	Decision conflict	Satisfaction	Other
	No difference		— — Better	Worse — —	No difference	No difference Persistence With choice
No difference	No difference	— More realistic	— Better clarity of values	— Better		More acceptable
— Trend, increased HRT			— Trend, better in those changing status quo (choosing HRT)	No difference		No difference Acceptability
	No difference			No overall difference Trend, better in uncertain patients		— More acceptable, no difference in anxiety
— — Increased vaccine			— — Better			
— Decreased trial entry		— More realistic				
No difference		More realistic —		No difference		Decreased symptoms and work loss —
No difference		No difference				No difference in acceptability
— Trend, decreased PSA		— More realistic				
— Decreased PSA						
— Decreased mastectomy	No difference					
— Trend, decreased mastectomy	— Trend increased					No difference in decision-making involvement
— Increased trial entry	No difference				No difference	

these differences and the cost-effectiveness of decision aids remain to be established.

Research Implications

Evaluation of Decision Aids

There are several gaps in research on decision aids. More research is needed on (*a*) how decision aids perform for different clinical decisions; (*b*) their acceptability to practitioners; (*c*) their acceptability to diverse patient groups; (*d*) their effect on patient–practitioner communication; (*e*) their downstream effects on persistence with the decision, distress, regret, and health-related quality of life; and (*f*) the optimal strategies for disseminating and for implementation.

Most evaluation studies are fraught with methodologic difficulties. They cannot be double-blind studies. Those studies that randomize patients rather than practitioners have contamination problems that narrow the differences that will be detected. Those studies that randomize practitioners need to be very large because of cluster sampling. Moreover, they may have selection biases because clinicians, knowing their assignment, may (*a*) be more or less enthusiastic about recruiting patients or (*b*) recruit different types of patients. Despite the researchers' best efforts, it is very difficult in a real-world setting to present the decision aid at the appropriate time to patients who are eligible to consider all of the options in the aid. Furthermore, efficacious interventions may have no effect if either patients or practitioners, or both, are extremely polarized toward one of the options at baseline. When postintervention measures are administered after the consequences of the choice are known, it is very difficult to avoid having the outcome color the patients' evaluation of satisfaction with the decision-making process and the decision.

We recommend that future studies should

- Examine the effect of decision aids on a broader range of decisions with a more comprehensive range of patient and practitioner outcomes;
- Select patients who are at the point of decision making for whom the choices in the aid are relevant;
- Measure patients' and practitioners' baseline predispositions toward the choices;
- Have sample sizes large enough to detect clinically meaningful differences in decisions among the undecided subgroup of patients;
- Measure patients' perceptions of practitioners' opinions;
- Have a usual care arm and describe clearly what usual care comprises; and
- Describe clearly what was in the decision aid and how it was used in the diagnostic or treatment trajectory.

Evaluation of Methods in Decision Aids

The incremental efficacy of including different strategies in decision aids should be explored. A starting point would be to evaluate the additional strategies outlined in the Cochrane definition *(11)*. Entwistle et al. *(18)* also outline several important issues regarding the presentational aspects of decision aids. Care should be taken in deciding which methods should be compared, considering the expense of an adequately powered study and the rather small differences observed in previous work. The use of nonpatient groups should be considered carefully. Although they make the study more feasible, it is difficult to generalize results to patients who are actually facing the decision and relying on more than the information in the aids to make their decisions.

We recommend the following:

1) The selection of strategies for evaluation should be based on whether the studies have the potential to exert a strong influence on decisions; vary considerably in their use in current decision aids; and contribute significantly to the cost, complexity, and time required to administer decision aids.
2) Clinically important differences should be defined *a priori*, and studies should be adequately powered.
3) Ideally, methods studies should include patients actually faced with the decisions and should evaluate cost-effectiveness.

Coordinating the Future Development of Decisions Aids

Decision aids have been developed on the basis of academic expertise and interest and sometimes evidence of population need. As the field matures, there is a need to focus more attention on systematic and standardized approaches to needs assessment, development, and evaluation. In Table 5, we have posed seven key questions that may be considered when deciding whether and how to develop a decision aid. These questions should be considered not only by individual research teams but also by cancer agencies with a system perspective. The order and depth of investigation of each question depend on the type of decision, the extent of previous research in the area, and the constraints and perspectives of the developers. The seven questions are as follows:

1) Is there a need for a decision aid? Needs assessment involves the compilation of evidence about the nature of the decision difficulty, the numbers affected, practice and preference variation, availability of aids elsewhere, and demand for the aid. Methods for needs assessment are varied, and data are obtained from primary or secondary sources or both. It is important that needs are defined from the perspective of potential users, both patients and practitioners.
2) Is it feasible to develop a decision aid? Feasibility is assessed to determine that the aid can be developed with available evidence and resources and can be delivered and updated in a timely, accessible, and acceptable manner.
3) What are the objectives of the decision aid? The objectives of the decision aid should be stated explicitly. Examples are identified in Table 5. The objectives drive the selection of the framework, intervention strategies, and evaluation methods.
4) Which framework will drive its development? Depending on the objectives, several frameworks are available to guide decision aid development *(8,12–19)*.
5) Which methods will be included in the decision aid? In selecting the decision support methods, the developer needs to determine how much emphasis will be placed on preparing the patient and the practitioner. The specific decision support methods, content, and delivery methods depend on the nature of the decision, the needs of the decision maker, the feasibility constraints, and the objectives of the decision aid.
6) Which designs and measures will be used to develop and evaluate the decision aid? Development and evaluation depend on the objectives of decision aids. Developers need to decide on the sampling and design architecture, the criteria for evaluation, and the measurement tools that will be used to operationalize the criteria.

A key issue is what are the primary criteria that should be used to evaluate efficacy? Examples of criteria currently in use are listed in Table 5. Entwistle et al. *(19)* provide an excellent overview of potential criteria depending on the model of patient involvement (shared decision making, individual informed choice, professional as agent for the patient, promotion of rational decision making, promotion of a particular choice). One of the biggest dilemmas is defining efficacy when choices depend on personal values for the outcomes. We maintain that decisions and outcomes of decisions should be evaluated from the perspective of the patients' values. Our assumptions are that 1) patients are unlikely to be able to value an option and communicate it to others unless they know what is involved and what outcomes are likely; and 2) once informed about options, patients are unlikely to implement or persist with an option that does not reflect personal values. Therefore, we consider a values-sensitive decision to be a good one if it is informed (based on adequate knowledge of options and outcomes and realistic expectations), based on personal values, and implemented. Moreover, the outcomes of a good decision should improve health outcomes (assessed using values-based measures) and persistence with choices (especially when the underlying reason for nonpersistence is a mismatch in values).

Another challenge is measuring the degree to which a decision is "consistent with personal values." This has been assessed with the use of the self-reports of patients *(68)*, multiple correlation coefficients between values and intentions *(12,67)*, percentage accuracy in discrimination between values and choice *(68)*, odds ratios from logistic regression predicting choices from preferences *(59)*, congruence between recommendations on the

basis of expected utilities and choices *(5)*, and agreement between choices and patients' treatment thresholds for median survival *(58)*. The relative advantages of these methods and others, such as preference-based health outcomes and quality-adjusted life years, should be explored.

7) How should the decision aid be disseminated? Dissemination involves the targeted distribution and promotion of the use of the decision aid. Six key elements of research transfer and use *(74)* are presented: potential adopters, practice environment, the evidence-based innovation (e.g., the decision aid), strategies for transferring the evidence into practice, evidence adoption, and outcomes. These elements are systematically monitored before, during, and after any research transfer efforts. The data generated by monitoring are used to 1) identify potential barriers and supports to research use associated with the potential adopters, the practice environment, and evidence-based innovation; 2) provide direction for selecting and tailoring transfer strategies; 3) track the progress of the transfer effort; and 4) assess the adoption of the evidence and its effect on outcomes of interest. Although dissemination is identified as a final step, it should be addressed early in the development process so that the aid is acceptable to potential users and has a greater potential for adoption. Therefore, dissemination questions can be posed during the needs and feasibility phases. Development and review panels can include potential users (practitioners and patients) and partners who may assist with dissemination (consumer groups, health professional organizations, disease foundations, and public education agencies).

Standards for Developing Future Decision Aids

The definition of a patient decision aid is open to broad interpretation, and materials of variable quality have been produced. Consumers expect to receive free health information and may have difficulty distinguishing the wheat from the "free" chaff unless certain standards are set in their development.

Many of the better decision aids have the following characteristics:

(*a*) They use evidence-based statements of benefits and risks from credible sources, refer to the quality and consistency of empirical studies, and use systematic overviews that extend shelf life and enhance updating.

(*b*) They are balanced in presenting all options (including doing nothing), benefits and risks, and, when included, examples of others' decisions and opinions.

(*c*) They identify the qualifications of the developers, including multidisciplinary expertise as evidence interpreters, communicators, practitioners, consumers, or disseminators, as well as conflicts of interest.

(*d*) They demonstrate commitment to update by 1) using expiry dates indicating the expected shelf life of the information, 2) mentioning upcoming trials that may shift policy, and 3) demonstrating linkage to an ongoing and credible evidence-analysis process (e.g., the Cochrane overview groups, the AHCPR PORTS or evidence centers, or the Cancer Care Ontario Practice Guidelines Initiative).

(*e*) They state the sources of funding in development, including potential conflicts of interest.

(*f*) They describe the efficacy of the decision aid in promoting evidence-informed choice, including acceptability, improvements in knowledge, etc.

Dissemination Issues

Table 6 describes how Lomas *(75)* applies his "research to policy" framework to shared decision-making programs. For each stage, from hypothesis generation to policy applications, he identifies key issues. Since the publication of the framework in November 1997, remarkable progress has been made in promoting shared decision making. Conceptual frameworks have been developed. Aids focused on cancer outcomes have been developed and pilot tested. Efficacy trials have been completed or are in progress. The results have been communicated broadly. Moreover, some of the aids are now being used by practitioners and patients.

One controversial issue is that of disseminating decision aids before large efficacy trials are completed with all relevant endpoints. Entwistle et al. *(18)* clarified the conflict by describing two rationales for promoting "evidence-informed patient choice" via mechanisms such as decision aids. The first is that one has a basic moral obligation to provide individuals with information and choice about their health care. Therefore, evidence-informed patient choice is the desired end. If decision aids can accomplish this end, as they have consistently demonstrated to date, then the tools should be disseminated. Much of the educational material available to patients has not been developed and tested as rigorously as decision aids. Moreover, some materials are developed by industries with vested interests in promoting increased use of their products. The public and nongovernmental organizations (NGOs) are clamoring for good decision support tools. NGOs and research institutes have sponsored the dissemination of decision aids that have been demonstrated to promote informed patient choice.

The second rationale is a consequentialist argument, based on the hypothesis that informed patient choice will lead to other beneficial outcomes. It is, therefore, a means to a desirable end, such as greater clinical effectiveness, health gain, individually appropriate utilization, reduced expenditures on inappropriate interventions, reduced litigation, and so forth. According to Entwistle et al. *(18)*, accepting the consequentialist argument means that we need to examine the benefits and harms of promoting evidence-informed patient choice across a whole range of health care decisions, patient groups, health care settings, and forms of decision support. Bernstein et al. *(49)* also argued for more studies of cost-effectiveness, given the expense of developing and disseminating these tools and the modest benefits observed in the trials to date. Therefore, proponents of this rationale would argue for more investigation before wide-scale dissemination.

A practical solution may be 1) to distribute aids that are needed, that are affordable, and that have been demonstrated to promote evidence-informed patient choice; 2) to continue to examine the downstream effects of decision aids in trials with usual care controls for new clinical decisions; and 3) to continue to explore (in methods studies) the most cost-effective strategies for achieving evidence-informed choice.

Conclusion

In this paper, the published evidence of the efficacy of decision aids has been presented. Although we know decision aids improve decision making, the downstream effect on persistence with decisions, health-related quality of life, and costs remain to be established. The implications for further research have been

Table 5. Developing and evaluating decision aids: questions and methods

Issue	Method (reference No.)
1. Is there a need for a decision aid?	
What are the decision-making needs of patients and practitioners?	Conduct key informant interviews, focus groups, or surveys to elicit patients' and practitioners' perceptions of decisions perceived as important and difficult, usual roles and decision-making practices, barriers and facilitators in providing or accessing decision support, potential strategies for overcoming barriers
What makes the decision difficult?	Review systematic overviews, decision analyses, and preference studies to determine whether benefits marginal or uncertain, risks material or uncertan, value trade-offs between benefits and risks, and variation in preferences, for outcomes
Are sufficient numbers affected and how are they affected?	Review databases, demographic or morbidity statistics, and population surveys
Is there sufficient variation in use?	Review practice atlases, utilization data, or practice variation studies
Are there decision aids available to meet these needs?	Review published overviews, reports; contact centers that produce aids
Is there a demand for decision aids and what methods are preferred?	Conduct market surveys *(81)*
2. Is it feasible to develop a decision aid?	
Are there adequate resources?	Assess finances, availability of experts with credibility, networks, and commitment to ongoing update; link to established overview and dissemination networks
Is there enough evidence of benefits and risks to incorporate into a decision aid?	Review systematic overviews with appraisals of the quality of evidence
How quickly is the evidence expected to change?	Review ongoing trials
Will aid be accessible or acceptable to users?	Conduct focus groups or market surveys
3. What are the objectives of the decision aid?	
Objectives focused on improved decision making	Improve knowledge of the clinical problem, options, outcomes, and variation in patient or practitioner opinions and practices. Create realistic expectations of outcomes, consistent with available evidence. Clarify personal values for outcomes and promote congruence between patients' values and choice. Reduce patients' and practitioners' decisional conflict (uncertainty) about the course of action to take. Promote implementation of choices. Improve patients' or practitioners' satisfaction with decision making.
Objectives focused on outcomes of decision	Promote patients' persistence with choice. Reduce patients' distress from consequences of decision. Improve patients' health-related quality of life. Promote informed use of resources by patients and practitioners.
4. Which framework will drive its development?	
	Charles et al. *(13)* distinguish shared decision making from other decision-making approaches. Entwistle et al. *(18,19)* define evidence-informed choice and outline different criteria for evaluations depending on objective. The Hershey et al. framework *(15)* developed for AHCPR has a health services and informatics perspective. The Llewellyn-Thomas framework *(14)* has a special focus on types of preferences and placement in sociopolitical context. Mulley *(16)* places shared decision making in the context of outcomes research. Ottawa Decision Support Framework *(12)* prepares practitioner and patient and has a clinical and behavioral focus. Rothert et al. *(8,17)* describe mutual roles of patients and practitioners in decision making; they focus on information and values.
5. Which methods will be included in the decision aid?	
Patient or client decision support	Information regarding options and outcomes • Content: clinical problem, options, outcomes • Detail in describing outcomes: define outcomes; describe physical, emotional, social effect; use narrative/scenario styles • Probabilities: none; numerical frequencies or percents *(7)*, graphic pie charts *(6)*, 100 people *(4,12)*, qualitative (low, moderate, high) • Tailored probabilities: not tailored; stratified by personal risk factors • Evidence for statements: references included or not Values clarification • Implicit *(4,6)* • Explicit methods such as weigh scale exercise *(7)*, treatment trade-off task *(82)*, relevance chart *(8)*, decisional balance sheets *(83)*, formal utility assessments *(5,9)* Information on others • None • Cases of different choices *(4,8,12)* • Statistics on variation in patients' decisions or practitioners' opinions Coaching or guidance in deliberation, communication, and implementation • Not included • Steps in weighing the benefits or risks *(8,12)* • Steps in discussing decision with a practitioner *(8,12)*

(Table continues)

Table 5 (continued). Developing and evaluating decision aids: questions and methods

Issue	Method (reference No.)
	• Tips on managing consequences of choices Delivery (personal counseling supplemented by) • Generic tools • Decision board *(6)* • Take home audio-guided workbook *(7)* • Interactive videodisc or linear video *(4)* • Computer-based tool *(70)* • Group lecture or workshop *(18)*
Practitioner decision support	Content: scientific evidence regarding decision, rationale for decision aid, efficacy of decision aid, timing and use in practice, scientific references Delivery: manual, video, lecture, workshop, hot line, academic detailing
6. Which designs and measures will be used to develop and evaluate decision aid?	
Development panel	Participants (researchers, clinicians, educators, patients, opinion leaders) Methods (iterations of drafts, feedback, revisions, feedback, etc.)
Review panels	Participants (potential users: practitioners, patients who have already made decisions) Methods (focus groups, personal interviews, questionnaires to elicit acceptability, etc.)
Pilot studies	Participants (patients at the point of decision making) Designs* XO XO: test–retest 2 weeks apart (stability of preferences and reliability of measures) OXO: before/after study uses baseline questionnaire, decision aid, or posttest questionnaire XO: posttest only, usually with preestablished criteria for success (e.g., 70% on knowledge test)
Trials	Participants (patients at point of decision making, practitioners) Designs* Quasi experiment OCO OXO Randomized trial (with unit of randomization either the patient or practitioner) ROXO ROCO
What are the criteria for evaluation?	
Knowledge	Knowledge or comprehension test *(4,7,8,47)*
Expectations of outcomes	Probability scales *(7,84–87)*, likelihood scales *(26)*
Clarity of values	Values subscale of Decisional Conflict Scale *(7,36,68)*
Agreement between choice and values	Statistical relationship between values and choices *(5,7,12,58,59,67,68)*
Realistic perceptions of others	Perceptions of percentage of practitioners or patients choosing options; subjective norms *(26)*
Decision	Choice question (option *x*, option *y*, unsure); choice predisposition *(7)*
Decisional conflict	Decisional conflict scales for patients *(7,36,88)* and providers *(89)*
Skill in decision making	Self-efficacy scale *(8,88)*, implementation data
Satisfaction with decision making	Decision satisfaction inventory *(4)*; satisfaction with decision *(8,90)*; satisfaction with preparation for decision making (O'Connor A, et al.: unpublished data)
Acceptability	Acceptability questionnaires—shared decision-making tool *(42–45)*, Ottawa tool *(68)*
Use of decision aid	Utilization data
Participation according to needs	Congruence between preferred and actual role in decision making *(91,92)*
Persistence with decisions	Survey of decision over time; implementation data
Reduced distress from outcomes	Condition-specific symptom and side effects distress scales
Health-related quality of life	Generic, condition-specific, preference-based
Use of resources	Analysis of utilization data
Costs	*See* the Hersey et al. *(15)* framework and the Nease and Owens *(94)* cost-effectiveness model
7. How should the decision aid be disseminated (74)?	
What are potential adopters: attitudes toward innovation or change; knowledge, attitudes, and skills to use aid; preference for shared decision making?	Conduct focus groups, key informant interviews, environmental scans, or surveys of potential users
What are the environmental barriers and supporters?	Conduct focus groups or surveys to identify the following factors: • Social (likely supporters and opposers; presence of opinion leaders as supporters; predominant belief system regarding shared decision making) • Structural (operational tools and processes, regulations, quality assurance criteria, to encourage or hinder use of aid; resources to support dissemination of aid) • Other incentives and disincentives
Will the evidence-based innovation meet expectations of target audience?	Conduct focus groups or surveys
Which transer strategies should be used?	Tailor strategies according to needs; diffusion strategies (advertisements, publications, or Internet); dissemination strategies (targeted mailings); implementation strategies (education programs or administrative changes)
Is the aid being adopted, and is it having the expected effect on outcomes?	Analyze databases; conduct quality assurance studies, surveys, and implementation studies to determine whether the aid is being used by the expected audience in the expected manner and whether it is having the expected effect on health and economic outcomes and evidence-based decision making among patients or clinicians

*C = control intervention; O = observation of effect; R = randomization; X = decision aid intervention.

Table 6. Stages of "Research to Policy" by Lomas *(75)* applied to shared patient decision making

Stages and issues	Progress to November 1998
Hypothesis generation—shared patient decision making is valued	Preferenee for shared decision making varies by patient characteristics and type of decisions *(77–79)* Patients desiring shared decision making do not always have the opportunity *(80)* Some evidence that oncologists endorse the concept *(81)*
Methods development—to elicit patient values and present risk–benefit information	Issues confirmed by AHCPR report *(15)* and Entwistle et al. *(18)* paper Several methods studies completed (Table 4)
Causal model or "theory"—lack of patient involvement leads to "unwanted interventions"	Several frameworks published; hypothesized effects are observed: help uncertain, base decisions on improved knowledge, realistic expectations, and personal values; effect on satisfaction variable; trend toward reduced preference for more intensive therapies
Study-specific evaluation (efficacy)—randomized trials under ideal controlled decisions	Seven trials show decision aids improve the quality of decisions and show trend in affecting decisions; several trials in progress in the United States, the United Kingdom, Canada, and The Netherlands
Knowledge summary and synthesis	Annotated bibliography *(40,41)*, AHCPR report *(15)*, and Cochrane systematic overview in progress *(11)*
Results communication—to providers or patients, media, journals, etc.	Conferences, workshops or journals; shared decision-making concept has ben discussed in media (e.g., National Press, magazines, local press, and television)
Application *to* policy world—what are the barriers to using decision aids under different health care funding and organizational arrangements?	Improving access: Canadian National Forum on Health identified the need to engage public in decision making regarding personal health and to improve access of providers and public to evidence-based information. Funding allocated to improve information systems. U.K. Kings Fund on Informed Choice launched. Some decision aids in hands of providers and consumers
Application *in* policy world—details of implementing shared patient decision making for specific jurisdiction, institution, or even practitioner	Practice policies: Canadian and U.S. practice guidelines beginning to identify role of patient preferences for some decisions (breast cancer surgery, hormone therapy). Canadian and U.S. nongovernmental agencies have endorsed some of the decision aids and are disseminating information on them Funding development, evaluation, updating, and dissemination: issue unresolved Standards for decision aids: issue unresolved: evidence-based, sponsorship, expiration dates Regulating/legislating use under certain conditions: several states have legislated that women with breast cancer should be offered mastectomy or lumpectomy options. Canada has a practice guideline. Institute for Clinical Evaluative Sciences atlas confirms practice variation in mastectomy or lumpectomy use. U.S. National Committee on Quality Assurance expands criteria to include quality of counseling on options that incorporates personal risk and preferences Service delivery: U.S. Foundation for Informed Medical Decision Making delivering decision support services for third-party payers

identified. There are outstanding issues regarding coordination, standards, and dissemination.

References

(1) Eddy DM. A manual for assessing health practices & designing practice policies the explicit approach. Philadelphia (PA): American College of Physicians; 1992.

(2) Guidelines for counseling postmenopausal women about preventive hormone therapy. American College of Physicians. Ann Intern Med 1992;117: 1038–41.

(3) Kassirer JP. Incorporating patients' preferences into medical decisions. N Engl J Med 1994;330:1895–6.

(4) Barry MJ, Cherkin DC, Chang Y, Fowler FJ, Skates S. A randomized trial of a multimedia shared decision-making program for men facing a treatment decision for benign prostatic hyperplasia. DMCO 1997;1:5–14.

(5) Clancy CM, Cebul RD, Williams SV. Guiding individual decisions: a randomized, controlled trial of decision analysis. Am J Med 1988;84: 283–8.

(6) Levine MN, Gafni A, Markham B, MacFarlane D. A bedside decision instrument to elicit a patient's preference concerning adjuvant chemotherapy for breast cancer. Ann Intern Med 1992;117:53–8.

(7) O'Connor AM, Tugwell P, Wells GA, Elmslie T, Jolly E, Hollingworth G, et al. Randomized trial of a portable, self-administered decision aid for postmenopausal women considering long-term preventive hormone therapy. Med Decis Making 1998;18:295–303.

(8) Rothert ML, Holmes-Rovner M, Rovner D, Kroll J, Breer L, Talarczyk G, et al. An educational intervention as decision support for menopausal women. Res Nurs Health 1997;20:377–87.

(9) Pauker SP, Pauker SG. The amniocentesis decision: ten years of decision analytic experience. Birth Defects Orig Artic Ser 1987;23:151–69.

(10) Llewellyn-Thomas HA, McGreal MJ, Thiel EC, Fines S, Erlichman C. Patients' willingness to enter clinieal trials: measuring the association with perceived benefit and preference for decision participation. Soc Sci Med 1991;32:35–42.

(11) O'Connor AM, Fiset V, Rostom A, Tetroe JM, Entwistle V, Llewellyn-Thomas HA, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Collaboration, Consumers and Communication Cochrane Review Group. Contact Paola Rio: rio@hna.ffh.vic.gov.au

(12) O'Connor AM, Tugwell P, Wells GA, Elmslie T, Jolly E, Hollingworth G, et al. A decision aid for women considering hormone therapy after menopause: decision support framework and evaluation. Patient Educ Couns 1998;33:267–79.

(13) Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997;44:681–92.

(14) Llewellyn-Thomas HA. Patients' health-care decision making: a framework for descriptive and experimental investigations. Med Decis Making 1995;15:101–6.

(15) Hersey J, Matheson J, Lohr K. Consumer health informatics and patient decision-making. Durham (NC): Agency for Health Care Policy and Research; 1997 AHCPR Pub. No.: 98-N001.

(16) Mulley A. Outcomes research: implications for policy and practice. In: Smith R, Delamother T, editors. Outcomes in clinical practice. London (U.K.): BMJ Publishing Group; 1995. p. 13–27.

(17) Rothert M, Talarcyzk GJ. Patient compliance and the decision making process of clinicians and patients. J Compliance Health Care 1987;2:55–71.

(18) Entwistle VA, Sheldon TA, Sowden A, Watt IS. Evidence-informed patient choice. Practical issues of involving patients in decisions about health care technologies. Int J Technol Assess Health Care 1998;14:212–25.

(19) Entwistle VA, Sowden AJ, Watt IS. Evaluating interventions to promote

patient involvement in decision-making: by what criteria should effectiveness be judged? J Health Serv Res Policy 1998;3:100–7.
(20) Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York (NY): John Wiley and Sons; 1976.
(21) Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981;211:453–8.
(22) Fischhoff B, Slovic P, Lichtenstein S. Knowing what you want: measuring labile values. In: Wallsten TS, editor. Cognitive processes in choice and decision behavior. Hillsdale (NJ): Lawrence Erlbaum Associates Inc.; 1980. p. 117–41.
(23) Janis IL, Mann L. Decision making: a psychological analysis of conflict, choice, and commitment. New York (NY): Free Press; 1977.
(24) Orem DE. Nursing: concepts of practice. 5th ed. Toronto (Canada): Mosby; 1995.
(25) Norbeck JS. Social support. Annu Rev Nurs Res 1988;6:85–109.
(26) Ajzen I, Fishbein M. Understanding attitudes and predicting social behaviour. Englewood Cliffs (NJ): Prentice Hall; 1980.
(27) Bandura A. Self-efficacy mechanism in human agency. Am Psychol 1982; 37:122–47.
(28) Greenfield S, Kaplan S, Ware JE Jr. Expanding patient involvement in care. Effects on patient outcomes. Ann Intern Med 1985;102:520–8.
(29) Greenfield S, Kaplan SH, Ware JE Jr, Yano EM, Frank HJ. Patients' participation in medical care: effects on blood sugar control and quality of life in diabetes. J Gen Intern Med 1988;3:448–57.
(30) Davison BJ, Degner LF. Empowerment of men newly diagnosed with prostate cancer. Cancer Nurs 1997;20:187–96.
(31) Frederikson LG, Bull PE. Evaluation of a patient education leaflet designed to improve communication in medical consultations. Patient Educ Couns 1995;25:51–7.
(32) Robinson EJ, Whitfield MJ. Improving the efficiency of patients' comprehension monitoring: a way of increasing patients' participation in general practice consultations. Soc Sci Med 1985;21:915–9.
(33) Butow PN, Dunn SM, Tattersall MH, Jones QJ. Patient participation in the cancer consultation: evaluation of a question prompt sheet. Ann Oncol 1994;5:199–204.
(34) McCann S, Weinman J. Empowering the patient in the consultation: a pilot study. Patient Educ Couns 1996;27:227–34.
(35) Roter DL. Patient participation in the patient-provider interaction: the effects of patient question asking on the quality of interaction, satisfaction and compliance. Health Educ Monogr 1977;5:281–315.
(36) O'Connor AM. Validation of a decisional conflict scale. Med Decis Making 1995;15:25–30.
(37) Haynes R, Taylor D, Sackett D. Compliance in health care. Baltimore (MD): Johns Hopkins University Press; 1979.
(38) Johnson JE, Fuller SS, Endress MP, Rice VH. Altering patients' responses to surgery: an extension and replication. Res Nurs Health 1978;1:111–21.
(39) Padilla GV, Grant MM, Rains BL, Hanson BC, Bergstrom N, Wong HL, et al. Distress reduction and the effects of preparatory teaching films and patient control. Res Nurs Health 1981;4:375–87.
(40) O'Connor AM. Consumer/patient decision support in the new millennium: where should our research take us? Can J Nurs Res 1997;29(3):7–19.
(41) O'Connor AM, Drake ER, Fiset VJ, Page J, Curtin D, Llewellyn-Thomas HA. Annotated bibliography: studies evaluating decision-support interventions for patients. Can J Nurs Res 1997;29(3):113–20.
(42) Spunt BS, Deyo RA, Taylor VM, Leek KM, Goldberg HI, Mulley AG. An interactive videodisc program for low back pain patients. Health Educ Res 1996;11:535–41.
(43) Deber R, Kraetschmer N, Trachtenberg J. Shared decision making: how does one measure success? [abstract]. Med Decis Making 1994;14:429.
(44) Liao L, Jollis JG, DeLong ER, Peterson ED, Morris KG, Mark DB. Impact of an interactive video on decision making of patients with ischemic heart disease. J Gen Intern Med 1996;11:373–6.
(45) Wagner EH, Barrett P, Barry MJ, Barlow W, Fowler FJ Jr. The effect of a shared decisionmaking program on rates of surgery for benign prostatic hyperplasia. Pilot results. Med Care 1995;33:767–70.
(46) Laupacis A, Hing M, O'Connor A, Biggs J, Hart R. Stroke prevention in atrial fibrillation III (SPAF 3) investigators. A randomized trial of an audiobooklet (AB) decision aid in patients with atrial fibrillation (AF) [abstract]. Paper presented at the 14th annual meeting of the International Society for Technology Assessment in Health Care (ISTAHC); 1998; Ottawa (Canada).
(47) Flood AB, Wennberg JE, Nease RF, Fowler FJ Jr, Ding J, Hynes LM. The importance of patient preference in the decision to screen for prostate cancer. Prostate Patient Outcomes Research Team. J Gen Intern Med 1996; 11:342–9.
(48) Morgan, Matthew. A randomized trial of the ischemic heart disease shared decision making program [thesis]. Toronto (Canada): Univ. of Toronto; 1997.
(49) Bernstein SJ, Skarupski KA, Grayson CE, Starling MR, Bates ER, Eagle KA. A randomized controlled trial of information-giving to patients referred for coronary angiography: effects on outcomes of care. Health Expectations 1998;1:50–61.
(50) Maisels MJ, Haynes B, Conrad S, Chez RA. Circumcision: the effect of information on parental decision making. Pediatrics 1983;71:453–5.
(51) Volk RJ, Spann SJ, Cass AR. A randomized comparative trial of shared decision making for prostate cancer screening: one year follow-up [abstract]. Med Decis Making 1998;18:477.
(52) Chapman GB, Elstein AS, Hughes KK. Effects of patient education on decisions about breast cancer treatments: a preliminary report. Med Decis Making 1995;15:231–9.
(53) Bredeson C. Use of a disease specific decision aid to determine the minimal clinically important difference between autologous bone marrow transplantation versus standard salvage therapy for patients with high risk low grade non-Hodgkin's lymphoma [thesis]. Ottawa (Canada): Univ. of Ottawa; 1997.
(54) Drake E, Engler-Todd L, O'Connor AM, Surh LC, Hunter A. Development and evaluation of a decision aid about prenatal testing for women of advanced maternal age. J Genetic Couns 1999;8(4):217–33.
(55) Fiset V. Evaluating the effectiveness of a decision aid for patients considering treatment options for stage-4 non-small-cell lung cancer [thesis]. Ottawa (Canada): Univ. of Ottawa; 1998.
(56) Man-Son-Hing M, Laupacis A, O'Connor A, Wells G, Lemelin J, Wood W, et al. Warfarin for atrial fibrillation. The patient's perspective. Arch Intern Med 1996;156:1841–8.
(57) Schonwetter RS, Walker RM, Kramer DR, Robinson BE. Resuscitation decision making in the elderly: the value of outcome data. J Gen Intern Med 1993;8:295–300.
(58) Brundage MD, Cosby RH, Feldman-Stewart D, Gregg R, Dixon P, Youssef Y, et al. A pilot study of a decision aid for patients with locally advanced non-small-cell lung cancer [abstract]. Med Decis Making 1998;18:483.
(59) Barry MJ, Fowler FJ Jr, Mulley AG Jr, Henderson JV Jr, Wennberg JE. Patient reactions to a program designed to facilitate patient participation in treatment decisions for benign prostatic hyperplasia. Med Care 1995;33: 771–82.
(60) Molenaar S, Sprangers MAG, Oosterveld P, deHaes JC. The evaluation of an interactive computer programme on treatment options in early stage breast cancer: preliminary results [abstract]. Med Decis Making 1998;18: 482.
(61) Gruppen LD, Margolin J, Wisdom K, Grum CM. Outcome bias and cognitive dissonance in evaluating treatment decisions. Acad Med 1994;69(10 Suppl):S57–9.
(62) Sawka CA, Goel V, Mahut CA, Taylor GA, Thiel EC, O'Connor AM. Development of a patient decision aid for choice of surgical treatment for breast cancer. Health Expectations 1998;1:23–36.
(63) Sebban C, Browman G, Gafni A, Norman G, Levine M, Assouline D, et al. Design and validation of a bedside decision instrument to elicit a patient's preference concerning allogenic bone marrow transplantation in chronic myeloid leukemia. Am J Hematol 1995;48:221–7.
(64) Dolan JG. Are patients capable of using the analytic hierarchy process and willing to use it to make clinical decisions? Med Decis Making 1995;15: 76–80.
(65) Fetting JH, Siminoff LA, Piantadosi S, Abeloff MD, Damron DJ, Sarsfield AM. Effect of patients' expectations of benefit with standard breast cancer adjuvant chemotherapy on participation in a randomized clinical trial: a clinical vignette study. J Clin Oncol 1990;8:1476–82.
(66) O'Connor AM, Pennie RA, Dales RE. Framing effects on expectations, decisions, and side effects experienced: the case of influenza immunization. J Clin Epidemiol 1996;49:1271–6.
(67) Holmes-Rovner M, Rovner D, Kroll J, Rothert M, Padonu G, Talarczyk G.

Does decision support make people more rational? [abstract]. Med Decis Making 1995;15:418.
(68) O'Connor AM, Wells GA, Tugwell P, Laupacis A, Elmslie T, Drake E. Effects of an 'explicit' values clarification exercise in a woman's decision aid regarding postmenopausal hormone therapy. Health Expectations 1999; 2:21–32.
(69) Wolf AM, Narsen JF, Wolf AM, Schorling JB. The impact of informed consent on patient interest in prostate-specific antigen screening. Arch Intern Med 1996;156:1333–6.
(70) Llewellyn-Thomas HA, Thiel EC, Sem FW, Woermke DE. Presenting clinical trial information: a comparison of methods. Patient Educ Couns 1995;25:97–107.
(71) Goel V, Sawka C, Thiel E, Gort E, O'Connor A. A randomized trial of a decision aid for breast cancer surgery [abstract]. Med Decis Making 1998; 18:482.
(72) Van Ruiswyk J, GriffinR, Schapira M. Informing patients about prostate cancer screening risks [abstract]. Med Decis Making 1995;15:421.
(73) Street RL Jr, Voigt B, Geyer C Jr, Manning T, Swanson GP. Increasing patient involvement in choosing treatment for early breast cancer. Cancer 1995;76:2275–85.
(74) Logan J, Graham I. Toward a comprehensive interdisciplinary model of health care research use. Sci Commun 1998;20:227–46.
(75) Lomas J. Improving research dissemination and uptake in the health sector: beyond the sound of one hand clapping, NcMaster University Centre for Health Economics and Policy Analysis; 1997 Nov. Policy Commentary C97-1.
(76) Dolan JG, Kerr JE. Colorectal cancer screening: a multicriteria decision analysis [abstract]. Med Decis Making 1998;18:488.
(77) Degner LF, Russell CA. Preferences for treatment control among adults with cancer. Res Nurs Health 1988;11:367–74.
(78) Blanchard CG, Labreque MS, Ruckdeschel JC, Blanchard EB. Information and decision-making preferences of hospitalized adult patients. Soc Sci Med 1988;27:1139–45.
(79) Sutherland HJ, Llewellyn-Thomas HA, Lockwood GA, Tritchler DL, Till JE. Cancer patients: their desire for information and participation in treatment decisions. J R Soc Med 1989;82:260–3.
(80) Degner LF, Kristjanson LJ, Bowman D, Sloan JA, Carriere KC, O'Neil J, et al. Information needs and decisional preferences in women with breast cancer. JAMA 1997;277:1485–92.
(81) O'Connor A, Llewellyn-Thomas HA, Sawka C, Pinfold SP, To T, Harrison DE. Physicians' opinions about decision aids for patients considering systemic adjuvant therapy for axillary-node negative breast cancer. Patient Educ Couns 1997;30:143–53.
(82) Llewellyn-Thomas HA. Investigating patients' preferences for different treatment options. Can J Nurs Res 1997;29(3):45–64.
(83) Pender NJ, Pender AR. Health promotion in nursing practice. 2nd ed. Norwalk (CT): Appleton & Lange; 1987. p. 159–83.
(84) O'Connor AM, Pennie RA. Reliability and validity of measures used to elicit health expectations, values, tradeoffs and intentions to be immunized for hepatitis B. J Clin Epidemiol 1995;48:255–62.
(85) O'Connor AM, Davies BL, Dulberg CS, Buhler PL, Nadon C, Hastings-McBride B, et al. Psychometric properties of a health risk attitude measures in predicting cessation among pregnant smokers. Med Care 1993;31: 658–62.
(86) Woloshin S, Schwartz LM, Byram SJ, Fischhoff B, Welch HG. Scales for assessing perceptions of event probability: a validation study [abstract]. Med Decis Making 1998;18:490.
(87) Palda VA, Llewellyn-Thomas HA, Mackenzie RG, Pritchard KI, Naylor CD. Breast cancer attitudes about rationing postlumpectomy radiation therapy: applicability of trade-off methods to policy-making. J Clin Oncol 1997;15:3192–200.
(88) Bunn H, O'Connor A. Validation of client decision-making instruments in the context of psychiatry. Can J Nurs Res 1996;28(3):13–27.
(89) Dolan JG, Markakis KM, Beckman HB, Gleeson ML. Further evaluation of the provider decision process assessment instrument (PDPAI): a process-based method for assessing the quality of health providers' decisions [abstract]. Med Decis Making 1996;16:465.
(90) Holmes-Rovner M, Kroll J, Schmitt N, Rovner DR, Breer L, Rothert ML, et al. Patient satisfaction with health care decisions. the satisfaction with decision scale. Med Decis Making 1996;16:58–64.
(91) Degner LF, Sloan JA, Venkatesh P. The Control Preferences Scale. Can J Nurs Res 1997;29(3):21–43.
(92) Strull WM, Lo B, Charles G. Do patients want to participate in medical decision making? JAMA 1984;252:2990–4.
(93) Whelan TJ, Levine MN, Gafni A, Lukka H, Mohide EA, Patel M, et al. Breast irradiation postlumpectomy: development and evaluation of a decision instrument. J Clin Oncol 1995;13:847–53.
(94) Nease RF Jr, Owens DK. A method for estimating the cost-effectiveness of incorporating patient preferences into practice guidelines. Med Decis Making 1994;14:382–92.

Note

A. M. O'Connor and I. D. Graham hold Ontario Ministry of Health Career Scientist Awards. A. Laupacis holds a Medical Research Council of Canada Scientist Award.

Population Risk, Actual Risk, Perceived Risk, and Cancer Control: a Discussion

Howard Leventhal, Kim Kelly, Elaine A. Leventhal

Given the difficulty of converting population-based estimates of cancer risk into precise statements of individual risk, it is not surprising that (*a*) individual differences in risk perception are at best poorly correlated to the best available determination of "actual risk" and to behaviors to prevent and detect and treat cancer, and (*b*) success in bringing perceived risk into line with actual risk has been limited. These inconsistencies are of concern because individual perceptions of risk are thought to be important motivators of action for the prevention and early detection and treatment of cancer. Following the reviewer's suggestion that risk perceptions are readily influenced by contextual factors, we suggest examining risk perception in a self-regulatory framework in which both risk judgments and motivated action are products of underlying representations of cancer and the self. Self-assessments of risk may access only a part of the data necessary for motivation, whereas motivation to sustain action calls on a larger number of concrete features of the database (symptoms, time loss, consequences). Studies of cancer risk perception can make a major contribution to our understanding of processes involved in self-appraisals and self-management to maximize well-being and to avoid catastrophic disease. [Monogr Natl Cancer Inst 1999;25:81–5]

How do we expect studies of risk perception to help with the control of cancer? To answer this question, it is useful to define the role of "risk" in the ongoing practice of cancer control. For the epidemiologist, risk refers to the probability or likelihood of the occurrence of cancer in specific populations within a given time frame *(1)*. The epidemiologist's definition of risk also involves the identification of factors associated with, and hopefully causally related to, different probabilities of the occurrence of cancer. The biologist searches for inherited and somatic (environmental) genetic factors responsible for failure to control cell division. The objective is to identify the mechanisms (i.e., molecular and physiological processes) underlying specific risk factors. Risk is defined at this level by identifying genetic markers for mutated genes or familial markers (family history) indicative of vulnerability to cancer. The clinician attempts to translate these different sources of information into a probability or risk statement for the individual patient by matching the factors in the patient's risk profile that overlap with the factors defining risk in both population and biological studies. Each approach represents efforts to define "actual" risk. At this time, neither the epidemiological nor the biological data alone or in combination are sufficient to answer the clinician's or patient's question: "Precisely what is the probability that (this specific patient/I) will contract a particular type of cancer at a given point in time, i.e., at 50, 60, or 70 years of age?" In short, there is considerable ambiguity in translating epidemiological and biological knowledge into estimates of actual risk for specific individuals in specific time frames, even with such individualistic risk indicators as those derived from Gail modeling *(2)*.

The presumed value of knowing actual risk is that it can facilitate the control of cancer by encouraging preventive action and early detection and treatment for individuals at high risk. Some routes for cancer control appear to bypass direct communication with at-risk individuals and focus on population risk. For example, establishing and enforcing public policies that reduce exposure to environmental carcinogens can reduce risk for entire populations, as can employment and work rules that require risk-reducing behaviors and penalize risky behaviors. Individual behavior, e.g., participation in screening and preventive actions, can also be influenced by mass media reports of cancer in public figures and by the decision of medical practitioners to introduce screening into annual examinations, the latter involving compliance in the absence of a personal decision. Each of these routes involves influencing behavior. Even efforts that do not directly target at-risk persons must affect behavior by influencing policy makers, and in our political system, this means influencing voters, their elected representatives, and the bureaucracy empowered with enforcing policy. In sum, while these routes change behavior of those at high risk, the means by which they do so, e.g., policy changes and integrating screening into annual medical examinations, do not require a change in perceived risk.

The importance of perceived risk, therefore, is its presumed significance as a motivator of behaviors to prevent, detect, and manage cancer in cases other than those involving public policy or medical or employment requirements for compliance. It is assumed that perceptions of risk are related to motivation to act and to action, and that increasing the match between perceived risk (beliefs) and actual risk (reality) will encourage individuals to initiate and maintain preventive and treatment behaviors at a level that is appropriate to their actual risk and its source. As the papers by McCaul and Tulloch *(3)* and Rothman and Kiviniemi *(4)* in this monograph so ably indicate, all is not well with this assumption. First, it is clear that it is no easier to bring perceived risk in line with estimates of actual risk than it is to extrapolate from epidemiological and biological data to the actual risk of cancer for a given person at a specific point in time. That creating a match of an estimated actual risk to perceived risk is difficult is attested to by data from genetic counseling in which substantial effort in one-on-one settings falls short of creating

Affiliations of authors: H. Leventhal, K. Kelly, Institute for Health, Health Care Policy and Aging Research, Department of Psychology, Rutgers University, New Brunswick, NJ; E. A. Leventhal, Department of Medicine, Robert Wood Johnson School of Medicine, University of Medicine and Dentistry of New Jersey, New Brunswick.

Correspondence to: Howard Leventhal, Ph.D., Rutgers State University of New Jersey, Institute for Health, 30 College Ave., New Brunswick, NJ 08901 (e-mail: howardl@rci.rutgers.edu).

See "Notes" following "References."

© Oxford University Press

perfect matches *(5)*. Second, only a weak relationship exists between perceived risk and behaviors to prevent cancers or to detect and to treat cancer. In our brief commentary, we address issues respecting the empirical support for these generalizations, and then we suggest a framework that can help us to examine both the factors responsible for the match or mismatch of perceived and actual risk and the discrepancy between perceived risk and preventive behaviors. The framework that we propose 1) indicates that the weak association between perceived risk and behavior is to be expected and 2) provides a path for dealing with this expectation in both research and practice.

Major Issues in Risk Perception

Weak Association Between Actual Risk, Perceived Risk, and Cancer-Control Behaviors

McCaul and Tulloch *(3)* review a previously published meta-analysis by McCaul et al. *(6)* that makes clear that the association between measures of perceived risk of cancer and screening behaviors was statistically significant but small at best: an effect size of $r = .16$ between measures of risk and mammography for the 19 studies considered. Family history of breast cancer and breast symptomatology fared better, showing moderate effect sizes for relationship to mammography, .33 for family history (19 studies) and .30 for breast problems, e.g., symptoms. Two sets of factors suggest that these modest effect sizes should be expected. First, the dependent variables under study, e.g., screening behaviors such as mammography and prostate-specific antigen (PSA) tests, and so forth, are not solely controlled by individual volition and, therefore, would not necessarily reflect the individual risk perceptions. As McCaul and Tulloch *(3)* suggest, people see doctors for routine, annual and work-related medical examinations and may be given or asked to take a sereening test that they do to fulfill their perceived obligation as a patient. As O'Connor et al. *(7)* indicate in this monograph, however, the rise of consumerism, the movement toward evidence-based medicine, and clinical trials reporting the outcomes of various cancer-control procedures may increase the pressure toward patient involvement in decision making. Indeed, patients may be surprised to find themselves under increased pressure to make decisions, such as decisions with regard to PSA testing, that were formerly left to their physicians. Individuals may be uncomfortable with this new responsibility and believe that they lack sufficient scientific knowledge to fill this decision-making role. The change may encourage sharing information with family members and friends leading to social inputs encouraging screening, once again bypassing risk perceptions. None of these factors, however, is stable. Thus, shifts in the cultural and medical-practice landscape can alter the balance between perceptions of risk and other factors with respect to specific screening and treatment behaviors.

In addition, personal experience with cancer-control procedures, such as mammography, PSA tests, colonoscopy, and the like, can affect how these practices are perceived and can influence the magnitude of their relationship to perceived cancer risk. Women who find mammography embarrassing, uncomfortable, and painful may be less likely to undergo screening. Furthermore, if women believe that some risk results from the available preventive, detection, and treatment options, e.g., belief that radiation from mammography increases risk, screening may be refused regardless of their perceived breast cancer risk. Whereas these beliefs may affect the relationship of perceived risk to practices recommended by medical practitioners, they may not affect individual readiness to engage in alternative or complementary practices. Thus, perceived risk may predict other, non-medically recommended actions. Investigators may be looking to the wrong outcomes when studying the behavioral consequences of perceived risk.

Second, on the independent variable side, Rothman and Kiviniemi *(4)* indicate risk perceptions are notoriously variable. Judgments of risk change depending on the response format, e.g., verbal, numerical, or ladder, and part of the variability arises because people have a variety of difficulties understanding numerical probabilities [e.g., *(8)*]. Furthermore, as Rothman and Kiviniemi *(4)* point out, the success of efforts to change perceived risk is poor. People do not understand differences between risk over a given time frame and risk as a cumulative estimate, and they treat identical risk ratios differently, depending on the numerical values in which they are expressed, e.g., 1/10 versus 10/100, the latter of which appears to convey greater risk. The variability of risk judgments reduces their value as predictors of subsequent, behavioral outcomes.

Bringing Risk Perceptions in Correspondence With Actual Risk

What can be said about the possibility of bringing risk judgments into line with actual risk and stabilizing these judgments to improve their relationship to selected, behavioral outcomes? The review by Rothman and Kiviniemi *(4)* in this monograph strongly suggests that risk judgments are highly sensitive to contextual factors; we suspect variability in response to context is highly dependent upon the participants. Many examples can be cited; e.g., people who are healthy and less well informed or have not committed to a decision are likely more influenced by context. Perceptions of risk are inflated when physically healthy recipients of risk information are encouraged to think about their own family history of disease and environmental exposures that cause risk *(9)*. Also, the greater the ease of recall of personally relevant risk-increasing factors, the greater the perceived risk, and the greater the ease of recall of risk-reducing factors, the lower the perceived risk *(10)*. Sensitivity to contextual factors increases the variability in risk judgments and lowers their power as predictors of behavior. The data suggest, however, that sensitivity to context varies across issues and populations. For example, the summary of decision aids by O'Connor et al. *(7)* in this monograph makes clear that more than 70% of patients "... who have a stated preference at baseline ..." are uninfluenced by the relatively intensive exposure to information involved in the use of decision aids for choice of treatment. By contrast, half of the undecided are influenced by information and decision aids. The degree to which contextual risk perceptions are more or less modifiable depends on the issue (death from eancer, air pollution, or car accident), the participant population (healthy, newly diagnosed, or advanced stages of disease), and the method of influence.

An Approach to Risk Perception for Research and Practice in Cancer Control

Uncritical acceptance of the two main points raised by the reviewers, i.e., the weak relationship of risk perceptions to behavior and the difficulty of bringing perceived risk in line with actual risk, could lead to premature rejection of risk appraisal

and behavioral approaches for controlling cancer. Indeed, the many factors that our reviewers identify as responsible for these two, general conclusions suggest that future research and practice may prove fruitless if they proceed on the assumption that risk perceptions are stable, have a causal relationship to overt behavior, and must match actual risk to be effective tools for cancer control. Rejecting both the stability and causal properties of perceived risk should not be interpreted to mean that risk judgments are invalid; perceived risk and readiness for action may indeed vary from situation to situation and moment to moment, but they may do so in understandable ways. The key issue is whether we can generate a model for understanding risk perceptions that will be usable for both research and practice.

Judgments as Self-Appraisals

If a woman is asked to evaluate her overall health status or risk of breast cancer, she can reflect on her family and personal medical history and her current physical and psychological function and can then make a self-appraisal as though she were an external observer *(11)*. She could report her judgment of her health status or perceived risk of breast cancer on a verbal or numerical scale. Poorly worded questions and unclear definition of scale points will lower the quality of these data. Reducing these "contaminants" to zero or removing the aforementioned contextual factors causing variability in risk perceptions will not change an essential fact: Self-reports of risk and other similar judgments may be absolutely fine predictors of health outcomes and have absolutely *no* causal relationship to behavior. Self-assessments of health on 5-point scales are excellent predictors of mortality over 5- and even 20-year time frames *(12)*, yet there is no evidence that these assessments are causes of health or risk behaviors, and there is no reason to assume that they cause mortality. Thus, we might identify a group of 65-year-old women who judge their risk of breast cancer in the decade ahead as greater than one in 20, higher than the population risk of women their age, and the future may confirm the correctness of their judgments. However, these judgments may have absolutely no implication or causal relationship to activities to control breast cancer, and only the superstitious would believe that these judgments cause breast cancer. Some of the observations on which these personal risk judgments are based may be reflections of causal variables; others may not (Fig. 1). For example, a high frequency of cancer in the family may reflect an underlying genetic cause; however, emotional distress may directly increase perceived risk but have no relationship to any of the processes causing breast cancer. In short, the observations affecting judgments of risk and health status may or may not be reflections of factors causing the outcomes to be predicted by the judgment *(13)*.

Judgments as Reflections of Motivation and as Causes of Motivation for Action

A covert, mental evaluation or an overt, expressed judgment of health status or risk can represent, therefore, an abstract assessment of the self that is stripped of motivational significance. Neither the covert appraisal of risk nor its overt expression has motivational power; motivation for self-protective action resides in the significance or meanings that underlie these appraisals. The belief system underlying motivation has at least three interrelated belief systems as illustrated in Fig. 1: (*a*) the representation (or beliefs) about the disease, i.e., its identity (label and symptoms), time line (age of occurrence), consequences, per-

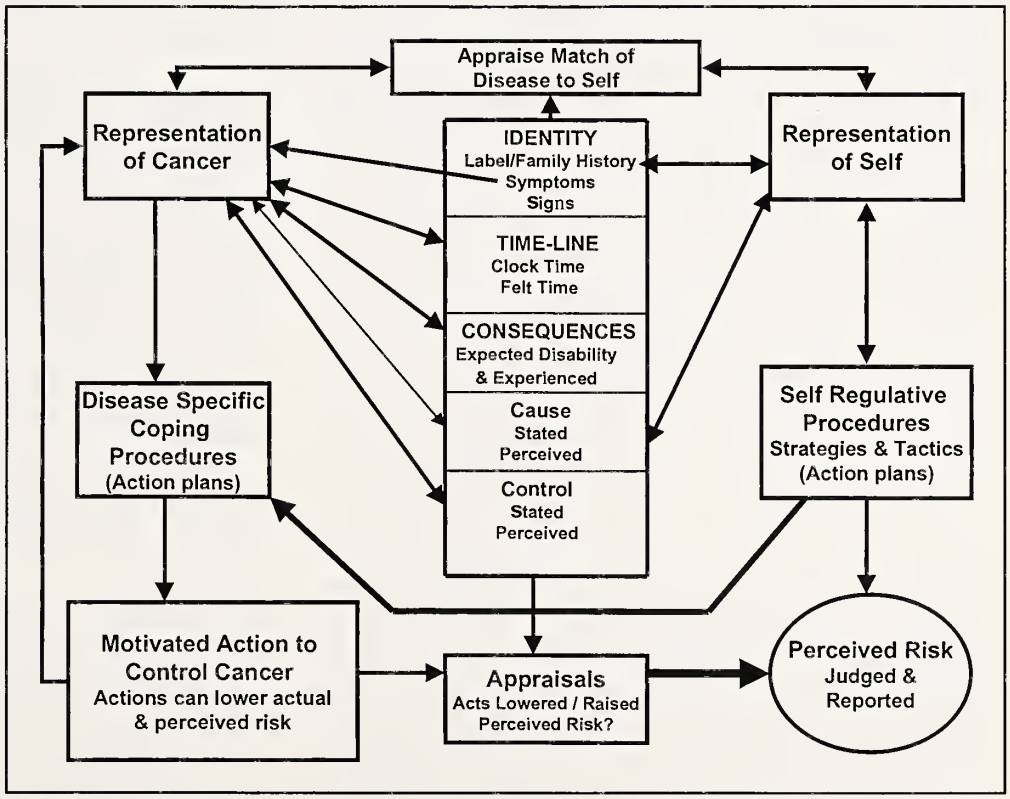


Fig 1. Common-sense perceptions of illness threats. Variables in five domains (identity, time line, consequences, cause, and control) define people's common-sense representations of illness threats. The overlap between the factors defining the representation of cancer (left side of diagram) and those defining the representation of the self (right side of diagram) establishes the self-relevance of the disease. The factors contributing to risk perception, e.g., family history of disease and beliefs about cause, may differ, however, from those needed to generate motivation, e.g., symptoms, time lines (time of disease onset within expected life span), consequences, and beliefs about control. Self-regulation procedures (e.g., general beliefs such as conservation of energy, reducing stress, diet, exercise, etc.) will maximize well-being, and longevity will affect perceived risk and moderate disease-specific coping procedures. Correspondence between perceived risk and motivated action to control cancer will increase with greater overlap between factors involved in the representation of cancer and the self, and this increase can result in a more congenial relationship between general, self-regulative procedures and medical recommendations for cancer control. For example, the general belief that "attacking problems early will facilitate positive outcomes" is more compatible with mammography and prostate-specific antigen testing than are beliefs that a vegetarian diet and avoidance of artificial substances, e.g., medications, are important for well-being and longevity.

ceived causes, and beliefs about its susceptibility to control *(14)*; (*b*) the representations of the self, e.g., self-identities, time line (expected years to mortality), causes (factors affecting vulnerability to disease), control (belief in ability to regulate environmental causes), and consequences (perceptions of changes in the self in coming years); and (*c*) disease-specific and self-relevant beliefs about the procedures for prevention, treatment, and recovery (e.g., risks and pain from surgery and radiation therapy). For example, a risk judgment may predict motivation to prevent and control cancer if the factors underlying it include observations of family history of disease and personal behaviors (high-fat diet, smoking) that indicate risk (i.e., the risk is personal), perceived consequences that are painful and distressing, causes that are controllable (i.e., the behavioral components are perceived as changeable), and a time line such that these factors are likely to cause cancer within the individual's lifetime. The representation of the illness and the relationship of this representation to beliefs about the self are the database for the risk judgment and the source of motivation for action. However, the factors necessary for forming a judgment of risk are not identical to those for creating motivation. Risk judgments can be based on family history of disease and the presence of somatic symptoms that create worry and feelings of vulnerability *(15,16)*. If other factors critical for motivation are absent, e.g., the time line for the disease and self (e.g., "Cancer won't strike before I die of heart disease"), consequences (e.g., "I can survive breast cancer if I get it"), and the relationship of its consequences to personal values (e.g., "It will not disrupt my family relationships"), judged risk will not be related to risk-reducing behavior.

Action is not solely a matter of the representation of the disease and the self; it is also affected by representations about available procedures for self-protection. Beliefs about the self and strategies for self-protection (self-regulatory procedures) are important moderators of action. People who believe that "I am what I eat" and "Toxins in my food may cause me to get cancer" will control intake to lower risk. If they believe stress causes cancer because, "Stress makes me feel ill," avoiding stress and conserving resources will be selected as routes to control risk. Action will be directed by these self-regulative strategies and associated, specific beliefs about particular procedural tactics for self-protection; e.g., green tea destroys toxins or reducing work commitments will lower stress. Medically recommended actions, e.g., early detection via mammography and treatment via surgery, will be subject to the same appraisal process. Overt action will take place if the procedures recommended for prevention are perceived as relevant to the perceived cause, as acting before the expected onset (time line) of the disease, and as doable within the framework of the individual's resources *(17)*, and if they fit the individual's generic, self-regulatory strategies.

A risk judgment unconnected with the set of factors necessary for motivated action will have but weak relationship to behavior unless specific efforts are made to link the judgment to those factors. In short, it is not enough to connect risk perceptions to actual risk: Actual and perceived risk must be connected to the tripartite system (i.e., representations of disease, self, and procedures) underlying behavior. Decision aids and counseling are designed to meet these objectives—i.e., to connect perceptions of disease and treatment risks and treatment outcomes to the self system. The degree to which counseling succeeds is likely to depend on the way in which it addresses the factors involved in the consequences, time frame, causes, control, and identity (symptoms, pain, etc.) of disease; the procedures for disease management; and the relationship of the two sets of beliefs to representations of the self, i.e., personal values and identities, time lines, self-regulatory strategies, etc., discussed in detail in this monograph by O'Connor et al. *(7)*. Furthermore, it is critical to remember that each of these sets of factors, e.g., somatic signs of presence of risk, time frames, consequences, and control, is multilevel. That is, each is represented in abstract, propositional forms of language and numbers, in concrete somatic sensations that stimulate fear of cancer and acquired images of family members and friends who suffered with cancer, and in yet more primitive or fundamental forms respecting the vigor of the physical self and its resistance to calamitous disease. The above conditions will ensure the formulation of a specific plan of action, e.g., a scheduled time for preventive action and a place for action, and adherence to a behavior recommended for cancer prevention.

The contributors to our symposium revealed the two major issues arising from studies of risk perceptions in the framework of cancer control: their weak relationship to behaviors to detect and control cancer and the difficulty of improving the correspondence of risk perceptions to actual risk, assuming actual risk can, in fact, be defined. In exposing the wide range of factors underlying these problems, our panelists exposed the contradictions in current data and, more importantly, the vigor and potential for increasing our understanding of risk perception and self-protective action by the theoretical and empirical analysis of perception of risk of cancer. Kurt Lewin is often quoted as saying, "Nothing is so practical as a good theory" *(18)*. The growing body of investigations on the behavioral aspects of cancer risk and cancer control suggests a variant of Lewin's quotation, i.e., "There is nothing so useful for good theory as the careful analysis of a practical problem."

References

(1) Ahlbom A, Norell S. Introduction to modern epidemiology. 2nd ed. Chestnut Hill (MA): Epidemiology Resources; 1990.

(2) McTiernan A, Gilligan MA, Redmond C. Assessing individual risk for breast cancer: risky business. J Clin Epidemiol 1997;50:547–56.

(3) McCaul KD, Tulloch HE. Cancer screening decisions. Monogr Natl Cancer Inst 1999;25:52–8.

(4) Rothman AJ, Kiviniemi MT. Treating people with health: an analysis and review of approaches to communicating health risk information. Monogr Natl Cancer Inst 1999;25:44–51.

(5) Kelly K, Leventhal H, Baran J, Schwalb M, Toppmeyer D, Much J, et al. Perceived risk and distress in individuals receiving counseling for BRCA1 and BRCA2 mutations. Poster session presented at the annual meeting of the Society of Behavioral Medicine; 1999 Mar; San Diego (CA).

(6) McCaul KD, Branstetter AD, Schroeder DM, Glasgow RE. What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. Health Psychol 1996;18:183–94.

(7) O'Connor AM, Fiset V, DeGrasse C, Graham ID, Evans W, Stacey D, et al. Decision aids for patients considering options affecting cancer outcomes: evidence of efficacy and policy implications. Monogr Natl Cancer Inst 1999;25:67–80.

(8) Shiloh S, Sagi M. Effect of framing on the perception of genetic recurrence risks. Am J Med Genet 1989;33:130–5.

(9) Weinstein N. Why it won't happen to me: perceptions of risk factors and susceptibility. Health Psychol 1984;3:431–57.

(10) Rothman AJ, Schwarz N. Constructing perceptions of vulnerability: personal relevance and the use of experiential information on health judgments. Pers Soc Psychol Bull 1998;24:1053–64.

(11) Bem DJ. Self-perception theory. Adv Exp Soc Psychol 1972:6:2–57.

(12) Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. J Health Soc Behav 1997;38:21–37.

(13) Benyamini Y, Leventhal E, Leventhal H. Self-assessments of health: what do people know that predicts their mortality. Res Aging 1999;21:477–500.
(14) Leventhal H, Leventhal EA, Contrada RJ. Self regulation, health, and behavior: a perceptual-cognitive approach. Psychol Health 1998;13:717–33.
(15) Easterling DE, Leventhal H. The contribution of concrete cognition to emotion: neutral symptoms as elicitors of worry about cancer. J Appl Psychol 1989;74:787–96.
(16) Cunningham LLC, Lauren LC, Andrykowski MA, Wilson JF, McGarth PC, Sloan DA, et al. Physical symptoms, distress, and breast cancer risk perceptions in women with benign breast problems. Health Psychol 1998; 17:371–5.
(17) Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psych Review 1977;84:191–215.
(18) Hovland CI, Janis IL, Kelley HH. Communication and persuasion. New Haven (CT): Yale University Press; 1953. p. 2.

Notes

Supported by Public Health Service grant AG03501 from the National Institute on Aging, National Institutes of Health, Department of Health and Human Services.

Matching Strength of Message to Strength of Evidence: a Discussion

Barnett S. Kramer

The four discussions in this session present different and sometimes contrasting aspects of cancer screening from the perspective of behavioral science. As a nonbehavioral scientist, I have tried to compare the behavioral aspects from the perspective of the public health professional in the field of cancer screening and to see the implications of a given behavioral strategy. I picked up, in essence, three general approaches or themes from the three papers on general population screening (Table 1). The genetic screening paper adds complexity that I will address at the end.

One *(1)* of the papers that we discussed today was primarily aimed at removing barriers to screening. For example, there was a statement in the paper that "the doctor should simply recommend screening." Some examples were given for mammography and prostate cancer screening, and McCaul and Tulloch in their presentation said some physicians urge prostate cancer screening to patients as the standard of care.

In their paper, O'Connor et al. *(2)* refer to this approach as "supplier induced demand." In essence, that is what happens to some patients during a routine checkup. The supplier virtually demands that the patient get a prostate-specific antigen test. We are seeing much of that in the medical practice today. The pharmaceutical or industrial suppliers are going straight to the public and advertising, or doctors are creating the demand for the types of things they do. The paper by Rothman and Kiviniemi *(3)* dealt with a conceptualized approach. Some of the means suggested to supply context were the provision of testimonials, imagery, and fear. More about that in a bit. The third paper O'Connor et al. *(2)* was a personal values-based approach or "informed choice" strategy, as they term it.

From my perspective, each of the three approaches has advantages and disadvantages. I listed attendant potential pros and cons in Table 2. The first approach is removing barriers. One of its advantages is that it is quite direct. You go for what you would like the patient's behavior to be. On the con side, the strategy can be directive. I think being direct has advantages but being directive has downsides, such as oversimplification of complex decisions or unproven interventions.

Table 1. Approaches to screening counseling

1. Remove barriers *(1)*
   - "First . . . simply recommend screening" (e.g., prostate)
   - Termed "supplier-induced demand" (2)
2. "Contextualized" approach *(3)*
   - Testimonials
   - Imagery
   - Fear
3. Personal values-based (2)
   - "Informed choice"

The contextual approach has a great deal of intuitive appeal. The word that comes to mind in terms of pros is "poignant." I could see how it would be extremely poignant and very compelling for someone to hear the context in which he or she is making the decision. In point of fact, we get a lot of contextual information these days. Congressional hearings are full of contextual approaches in which you will often, if not universally, have a panel of people who testify to very personal and moving experiences while a given committee is contemplating legislation on national policy. The recent Internal Revenue Service (IRS) hearings are an example. A panel of people gave stories of personal sufferings in dealings with the IRS. That is extremely poignant because most people will resonate with the suffering and fear. The problem is one of numerator versus denominator. You do not know how large the numerator of suffering people is. The denominator is many, in this case more than 100 million people. The question is: How reflective is the contextual story (i.e., numerator) of the general? This also has happened in hearings that have dealt with a variety of medical issues. You have someone who is personally affected. You have seen on national television people lamenting what happened to them within the medical system. And although it is poignant, it is hard to know whether it is oversimplified or if the problem is general. It is hard to know whether it is an anecdote and, sometimes, it is hard to verify its accuracy in the first place. In other words, accuracy of inference may suffer at the altar of poignancy.

Finally, informed choice has the pros of empowerment in engaging the patient in an informed decision. It does have its downsides as well. It can be time consuming for the health professional. It can be quite confusing to the person trying to choose among options. And it can affect compliance. If you, as a health provider or health professional, think you know exactly what the right thing is to do, and you complicate the message

Affiliation of author: Division of Cancer Prevention, National Cancer Institutes, Bethesda, MD.
Correspondence to: B. S. Kramer, M.D., M.P.H., National Institutes of Health, Bldg. 31, Rm. 10A46, MSC 2580, Bethesda, MD 20892.

Table 2. Pros and cons of suggested approaches to screening counseling

	Pros	Cons
1) Remove barriers	Direct	Directive
2) Contextual	Poignant Anecdotal	Oversimplified
3) Informed choice	Empowering	Time consuming Confusing Compliance

with any issues of pros and cons, the person may not comply and you might consider that a defeat of your purpose. Resources required for presenting complex messages are another problem.

An Additional "Axis" of Complexity

I have some additional thoughts in terms of research directions in behavioral approaches to cancer screening. The papers in this session deal almost exclusively with what may happen to an individual, whether medically or socially, in the screening setting. However, I think that there is yet another aspect to address in health communications to the individual (or to the public)—that is, not just what the range of outcomes is and the likelihood that these outcomes could occur, but the *strength of the evidence* that informs the discussion in the first place. The Physician Data Query (PDQ) of the National Cancer Institute provides a means to do this. Formal levels of evidence are attached to each discussion. This represents a completely different conceptual axis because it does not represent the actual health outcomes. It is the strength of data that support the statement about outcomes—good or bad. There can be strong, intermediate, or weak levels of evidence.

Table 3 collapses the five-level rating system of PDQ into three general levels. The strongest level of evidence comes from a randomized, controlled prospective trial, in which the only thing that has been changed is the actual intervention. The observed outcomes, therefore, reflect directly a cause-and-effect relationship between intervention and outcomes. Intermediate-strength evidence can come from a variety of study designs. The example in Table 3 is evidence from observational studies. These designs can suggest causation, but confidence is not nearly as strong. Rather than *causation,* they show *association* between intervention and outcome. The weakest level of evidence would be opinion, logic, and clinical judgment. In the area of screening, there often tends to be more opinion rather than other forms of evidence. The important bottom line of this second conceptual axis is to match the strength of opinion to the strength of the evidence, which I think is a somewhat new dimension that is often lost in crafting health-related messages for the public.

Table 3. Added component to informed choice in screening

Levels of evidence (PDQ*)
- Strong (e.g., clinical trials, controlled)
- Intermediate (e.g., observational studies)
- Weak (e.g., opinion, clinical judgment)

Key: Match strength of opinion to strength of message

*Modified from the NCI Physician Data Query (PDQ).

In summary, the first dimension of communication on cancer screening is the medical, social, and economic outcomes; both pros and cons; and the consequences of intervening, or not, medically. The other dimension is the actual strength of the evidence. There are even higher orders of complexity (Table 4). There may be multiple risks with which you have to deal. As discussed earlier in this conference, there is very little research going on in how to deal with communicating multiple risks. There are also multiple potential screening interventions, so one not only has to weigh risks versus benefits of a single screening test but also has to consider, for example, "Should I get a mammogram or the heart scan that's also advertised a lot," or "Should I get a barium enema or a sigmoidoscopy," and it goes on. (It is virtually impossible to get all medical screening tests advertised to the public.) One way to approach such a situation of multiple choices and trade-offs is to examine the number needed to treat. There's little information on its use as a communication tool, but research could be helpful. When a health professional is explaining a medical test or treatment, it may be, for example, helpful to say, "We would have to treat 10 people with your blood pressure level to avert one stroke," or "We would have to screen 2000 women your age for breast cancer annually for 10 years to increase life by 1 year," thus providing context among medical interventions.

There is also the complexity of communicating consequences of multiple tests, not just each one in isolation. The public is deluged by health messages. Apart from the strength of evidence, and totally apart from the actual outcomes, there are many screening tests of varying accuracy. Multiple professional societies issue numerous screening recommendations in isolation. What could be the consequences of following each and every recommendation? It has been estimated that about 125 screening recommendations are made to the U.S. public by various professional and advocacy groups. If each recommended test were independent and had a specificity of 0.95 (considerably better than most such tests), the likelihood that a perfectly healthy person would be declared healthy on all screening tests is 0.0016 (0.95^{125}). Even at the extremely unrealistic specificity of 0.99, the likelihood would be 0.28. That is a behavioral research challenge—not just approaching one screening tool at a time or one intervention at a time, but putting it into the full context of other recommendations and health strategies.

Can we help the person considering the messages with additional information such as level of evidence or comparisons to other strategies in terms of "number needed to treat"? Research on how to frame such messages seems important to me as a nonbehavioral scientist.

Finally, a comment on the presentation by Croyle and Lerman *(4)* on the more restricted area of genetic screening. We have gotten into a brave new world in which risk is not homogenous across the nation. As we learn about genetic tests, we are

Table 4. Dimensions of the message

1. Medical, Social outcomes
2. Strength of Evidence

Higher orders of complexity:
- Multiple risks
- Comparing multiple strategies: "Number needed to treat"
- Multiple tests: specificity problems
- Heterogeneous risk (e.g. genetics)

going to learn that there are subpopulations and founder populations who are at much higher risks. Then there are going to be the far larger number of people who have polymorphisms who are at moderately elevated risk. Almost everyone will have, in some sense, a "pre-existing condition," predisposing them to harms from specific exposures. If an individual is exposed to tobacco smoke or an industrial carcinogen, risk depends on how well he or she metabolizes and excretes carcinogens and procarcinogens or repairs molecular damage. The message may have to be far more complex and tailored than any that we have imagined before. We have immense and growing challenges in the field of risk communication.

References

(1) McCaul KD, Tulloch HE. Cancer screening decisions. Monogr Natl Cancer Inst 1999;25:52–8.

(2) O'Connor AM, Fiset V, DeGrasse C, Graham ID, Evans W, Stacey D, et al. Decision aids for patients considering options affecting cancer outcomes: evidence of efficacy and policy implications. Monogr Natl Cancer Inst 1999;25:67–80.

(3) Rothman AJ, Kiviniemi MT. Treating people with information: an analysis and review of approaches to communicating health risk information. Monogr Natl Cancer Inst 1999;25:44–51.

(4) Croyle R, Lerman C. Risk communication in genetic testing for cancer susceptibility. Monogr Natl Cancer Inst 1999;25:59–66.

Introduction of Section: Persuasion for the Purpose of Cancer Risk Reduction: Understanding Responses to Risk Communications

Lisa G. Aspinwall

Risk behaviors and responses to risk communications are complex and multifaceted. Two target articles *(1,2)* conclude that little longitudinal evidence shows that risk perceptions predict precautionary behaviors. This paper focuses on several questions raised by these perplexing findings that have implications for future research on risk communications. A pressing need exists to understand how people process risk information over time and how such processing may differ as a function of risk status, individual differences, social context, and other factors. I will review evidence and methods from the study of persuasion and attitude change that suggest several ways to study message processing to understand what kinds of thoughts are brought to mind following a persuasive communication, as well as how such thoughts may be related to subsequent beliefs and behaviors as people encounter new information and make risk-relevant choices. [Monogr Natl Cancer Inst 1999;25:88–93]

Exposure to a risk communication or the perception of some internal cue to action, such as a physical symptom, is supposed to heighten perceived risk of illness, which in turn prompts precautionary behavior. Increasingly, researchers have recognized that human behavior rarely conforms to this ideal sequence. First, information that one is at risk for serious illness may often be processed in ways that downplay one's personal risk. Second, the acknowledgement of high personal risk is no guarantee that one will enact precautionary behavior. In particular, the two target articles yield the same conclusion: the evidence linking high-risk perceptions to increased precautionary behavior is limited, both for behavioral risk-reduction measures *(1)* and for some kinds of cancer screening *(2)*. Instead, risk perceptions may be unrelated to subsequent behavior change and may even predict decreased precautionary behavior. In particular, for behaviors that are complex and for illnesses that are extremely severe or threatening, the proportion of studies showing a relation of risk perceptions to precautionary behavior is quite low *(3,4)*.

How and why do responses to risk communications deviate from the ideal sequence outlined above, and when are such deviations problematic? I will examine some reasons why the literature produces such mixed results concerning (*a*) the relation of risk communications to risk perceptions and (*b*) the relation of risk perceptions to precautionary behavior. I start with a brief review of contemporary approaches to understanding persuasion and attitude change that may be useful in developing and testing persuasive messages for cancer risk reduction and screening. I will next discuss the importance of understanding how people high and low in different risk factors for cancer may respond differently to risk communications, with implications for subsequent precautionary behavior. I will conclude with recommendations to focus on understanding a wider set of responses to risk communications and to examine personal, social, and experiential influences on such responses over time.

Contemporary Approaches to Persuasion and Attitude Change

Persuasion and attitude change have been central topics in social psychology since World War II *(5–7)*. Early studies focused on what came to be known as the "who says what to whom" paradigm of the Yale School; i.e., researchers examined the source of a persuasive message (e.g., expertise, trustworthiness), the message itself (e.g., one-sided versus two-sided arguments), and characteristics of the audience (e.g., education, prior knowledge) as determinants of successful persuasion. Contemporary approaches to attitude change, while they consider such factors, focus mostly on additional determinants of the audience's responses to persuasive communications—not only who says what to whom but also with what effect? Three factors have been found to be especially important in understanding how a target audience or individual will process a message and, accordingly, whether the audience will adopt the attitude advocated in the message: (*a*) motivation to process the persuasive message, (*b*) ability to process the persuasive message, and (*c*) cognitive responses to the message.

Motivation and Ability to Process the Persuasive Message

The most-studied factor in determining motivation to process messages carefully and extensively is self-relevance or personal involvement *(6)*. In experiments that vary the personal relevance of the message (e.g., by having college students read persuasive communications about requiring comprehensive examinations for graduation that will apply to their university or to another university), it is clear that conditions of high personal relevance create greater attention to message content. Highly involved audiences, among other things, are more sensitive to argument quality, responding favorably only to high-quality arguments and rejecting weak, low-quality arguments, even when they are provided by experts. In contrast, less involved audiences devote less effort to processing the message itself and may base their attitudes on factors other than argument quality, such as the expertise or attractiveness of the source. This distinction between more and less effortful processing has also been termed systematic versus heuristic processing *(5)*. Attitudes that are based on careful, systematic processing of a message are more resistant to persuasion than those that are based on less careful examination of the message.

Ironically, when the information in question is highly self-

Correspondence to: Lisa G. Aspinwall, Ph.D., Department of Psychology, University of Maryland, College Park, MD 20742–4411.

See "Note" following "References."

© Oxford University Press

relevant and negative (as is typical of cancer risk communications), it is often the case that those factors that increase motivation to process the message (e.g., factors that highlight the personal relevance of the risk information) may simultaneously decrease people's ability to attend closely to it. For example, Jepson and Chaiken *(8)* found that people with a chronic fear of cancer were less likely to detect logical errors in cancer-related messages than people without such fears. Messages designed to highlight personal vulnerability may create fear, anxiety, and distraction, which, in turn, have been shown to compromise information processing. These factors may lead people to process messages incompletely or to engage in biased processing to reduce their perceived risk. Researchers have known for some time that fear-arousing communications must be paired with information about offsetting or reducing the threat to be effective *(9)*. Other factors, such as time pressure and message complexity, have also been found to reduce people's ability to process persuasive messages.

Cognitive Responses to Persuasion

The third factor, which is highly influenced by one's motivation and ability to process a message, as well as one's prior knowledge, is cognitive responses to persuasion. Such responses are typically assessed by asking people to list all of the thoughts that come to mind following exposure to a persuasive communication and then coding these thoughts for agreement or disagreement with various aspects of the message. According to the elaboration likelihood model of persuasion of Petty and Cacioppo *(6)*, the success of a persuasive message depends on the thoughts the audience brings to mind following exposure to the message. If the audience generates thoughts agreeing with the message (termed cognitive elaboration), persuasion is more likely; however, if the audience generates thoughts disagreeing with the message (termed counterarguing), the audience is unlikely to be persuaded by the message and is also less likely to be persuaded in the future.

Interventions based on the counterarguing concept have been highly successful. In one study *(10)*, junior high school students taught to develop and rehearse counterarguments to advertisements and to peer pressure to smoke were much less likely to initiate smoking than students at a control school. Such results highlight the importance of understanding the thoughts that are brought to mind as people are exposed to risk-related communications and as they encounter risk-related choices in their daily lives as determinants of future attitudes and behaviors.

Unfortunately, it is a common finding that people readily generate counterarguments to communications suggesting that they are at high risk for cancer and other illnesses. One particularly common method of counterarguing involves using a higher standard for evidence that is preference inconsistent (e.g., bad news about the consequences of one's own behavior or information that suggests one's health risk is high) than for evidence that is preference consistent (good news about the consequences of one's own behavior or information that suggests one's health risk is low) *(11,12)*. For example, Liberman and Chaiken *(13)* exposed women who were frequent coffee drinkers to medical articles examining the link between caffeine and fibrocystic breast disease (FBD). Women who were frequent coffee drinkers were much more critical of the studies supporting the link between caffeine and FBD than women who were not frequent coffee drinkers, but they did not apply this same level of scrutiny to articles refuting the dangers of caffeine consumption. Evidence from recent National Cancer Institute (NCI) focus groups *(14)* is also consistent with these findings, suggesting that people are aware that they counterargue risk communications and that they require a higher standard of evidence to accept bad news about their risk.

From the elaboration likelihood theory of Petty and Cacioppo *(6)*, we can predict that such counterarguing—the generation of disagreeing thoughts in response to risk communications—will make people more resistant to persuasion by future risk communications because disagreeing thoughts may be more easily brought to mind with each successive exposure. The challenge to researchers and practitioners, then, is to understand the conditions that are most likely to trigger elaboration versus counterarguing of risk communications and to change the content or delivery of risk communications to reduce counterarguing and to increase message elaboration. In the following sections, I will discuss the specific research and policy recommendations made in the target articles and suggest some additional ways to study people's responses to risk communications in light of these challenges.

Recommendations for Research on Risk Communication

Test Responses to Risk Communications as a Function of Risk Status

My first recommendation is to test cognitive responses to a particular risk communication before using it in an intervention or media campaign. If the message triggers counterarguing in any segment of the audience, that segment may become more resistant to persuasion. Knowing the specific disagreeing thoughts that people may generate would allow researchers and practitioners to build these thoughts into the message; i.e., if we knew what kinds of disagreements were common, we could forestall these disagreements by addressing them explicitly in the persuasive communication. Messages that trigger large numbers of disagreeing thoughts should be reconsidered or abandoned, since they may reinforce the very cognitions the message is designed to change.

In many respects, this recommendation resembles an initial strategy of elicitation research in which researchers use open-ended measures, focus groups, and other measures to understand the spontaneously accessible information, motivation, and behavioral skills related to precautionary behaviors in specific risk groups *(15)*. More important, however, the present recommendation suggests that an additional step may be necessary. In addition to trying to understand what people know about their cancer risk to provide them with the information that they need, it is important to understand how they will react when such information is provided.

These tests of responses to risk communications should be conducted separately in different risk groups. A pressing need exists to understand how people who are high and low in actual risk, by virtue of their family history of cancer, their own health behavior, or other factors, respond to risk communications. As Vernon *(2)* points out, although studies have addressed fears surrounding and barriers to obtaining a mammogram, very little is known about the predictors of mammography among women at high risk for breast cancer. Both motivation and ability to process risk communications and the specific kinds of counter-

arguments generated in response to the message are likely to be markedly different for people with different risk factors. In the high-risk groups, the message is more likely to arouse fear and concern; therefore, adding materials to increase perceived vulnerability to cancer is likely to be counterproductive in that people already tend to know that they are at high risk. What people may want to know is how to reduce their risk or how to profit from early detection if their risk cannot be reduced. Likewise, emerging data concerning the prevalence of a pessimistic bias among people in high-risk groups (i.e., those who overestimate their risk) suggest that efforts to increase risk perceptions are not always necessary but that communications to address people's fears and to provide information about risk-reducing behaviors are. Studies of responses to risk communications about mammography and breast self-examination among high-risk women could tell us a great deal about how people manage the emotional consequences of their high-risk status and how such efforts are related to beliefs about screening and self-examination. In contrast, in low-risk groups, the goal of the communication may be to increase personal vulnerability but only to the level required for people to see the message as personally relevant.

Such studies may also provide valuable information about why the predicted relation of risk perceptions to precautionary behavior is so elusive—this relation may be different among high- and low-risk people. Specifically, high perceptions of risk may interfere with precautionary behaviors, and low risk perceptions may not necessarily preclude precautionary behaviors. For these reasons, the goal of risk communication should not necessarily be to make people appreciate their risk of serious illness but instead to increase risk-relevant knowledge, adherence to risk-reduction behaviors, and compliance with screening recommendations in ways that will improve their health and reduce their chances of illness.

As noted earlier, the relation of risk perceptions to precautionary behavior is especially weak for highly severe illnesses. One approach that has been successful in increasing precautionary behaviors for other life-threatening illnesses, such as human immunodeficiency virus (HIV) infection, has been to present precautionary behaviors in the context of less serious and less stigmatized risks. For example, in an intervention to reduce young women's risk of contracting HIV, Bryan et al. *(16)* geared the intervention toward reduction of sexually transmitted diseases (STDs) in general instead of HIV infection specifically. Of course, precautions that reduce one's vulnerability to STDs in general also reduce one's risk of HIV infection, but the recruiting materials and risk communications do not have to work against a strong negative emotional response to HIV or acquired immunodeficiency syndrome to be effective. People engaged in managing a less severe risk may be more able to process risk-relevant information and less motivated to engage in counterarguing. In this way, people may learn and practice risk-reduction behaviors without having to acknowledge their risk of a life-threatening illness. With respect to cancer risks, communications targeted to less threatening health concerns, but that still address the same behavioral risks (diet, smoking, etc.), may be more successful than those that emphasize cancer vulnerability, especially among people at high risk. Similarly, cancer screening might be presented as part of an age-appropriate overall health assessment rather than as a particular set of tests exclusively for cancer.

Examine Individual Differences That May Influence Responses to Risk Communication

In addition to being negative and threatening, cancer risk communications often have a second set of hurdles to overcome. As Gerrard et al. *(1)* point out, communications about behavioral risk reduction may also present information that reflects poorly on the self, namely, that one's own behavior is risky or foolish. Gerrard et al. suggest that such information is likely to trigger defensive or reactive responses, especially among people with high self-esteem. On the basis of studies of smoking cessation and sexual risk behavior, they suggest that people with high self-esteem are less likely to acknowledge that their own behavior increases their risk.

These findings suggest that some people's motivation to counterargue risk communications is especially high. Therefore, specific risk communications must be more sensitive to the possibility that a message may promote counterarguing, not only to reduce perceptions of personal vulnerability but also to save face, either to oneself or to others. Increasing evidence *(17)* suggests that many health-risk behaviors, such as excessive dieting, failure to use sunscreen, and failure to wear such protective devices as helmets or life jackets, are related to self-presentational concerns. In addition to being right, people wish to appear "cool," attractive, strong, and not overly concerned with risks, and such goals may compete against goals to prevent illness. For these reasons, continued efforts to present precautionary behaviors as more socially acceptable and positively, instead of negatively, self-defining are especially important.

The findings by Gerrard et al. *(1)* concerning self-esteem suggest that another important direction for future research will be to examine how people respond to negative information about themselves and to the conditions that may help people to process and elaborate, rather than counterargue, such information. In an experimental test of this question *(18)*, we found that giving people a boost in one area of the self-concept unrelated to health behavior before exposure to information about health risks promoted more open, less biased processing of risk-relevant information. Specifically, women who were frequent caffeine consumers were randomly assigned to complete a 10-item survey about their prior acts of kindness and generosity or to a neutral survey before their exposure to passages confirming or disconfirming the link between caffeine and FBD. The kindness survey was constructed so that all participants would receive high scores. Compared with women in the neutral condition, women who completed the kindness affirmation oriented more quickly to the information suggesting that caffeine was linked to FBD, reported greater beliefs in the link between caffeine and FBD, and reported greater perceived control over reducing their caffeine consumption. It is interesting that they also reported lower intentions to reduce their caffeine consumption at the end of the first session. However, at a 1-week follow-up, women who completed the affirmation were less likely to recall risk-disconfirming information than women in the neutral condition. Put differently, women who did not complete the self-affirmation showed the typical pattern of enhanced recall for information suggesting that their behavior was not risky, but this tendency was eliminated among women in the self-affirmation condition.

This particular self-affirmation method has not yet been tested extensively in the health domain, but several studies in

other domains reviewed by Aspinwall *(19)* similarly suggest that boosting people's sense of important self-beliefs, values, or competence in one area may help them manage self-relevant negative information in a second area. These findings suggest that it may be possible to reduce the tendency of people high in self-esteem to counterargue risk information in one domain by reminding them of their triumphs or strengths in other domains. Such affirmations, when carefully constructed so that all participants receive high scores, may also increase the ability of people with low self-esteem to manage negative information about themselves.

Understand Responses to Communications About Different Risk Factors for Cancer

Understanding how people process information that suggests that their own behavior places them at risk for illness is important, but it may be just as important to study how people respond to information that suggests that there is something else about them—their genetic makeup, their family history, or previous environmental exposure—that increases their risk but cannot be directly modified. The results of recent NCI focus groups *(14)* suggest an acute need to understand how people think about genetic risk (e.g., is there an inevitable one-to-one relation of genetic risk to the development of cancer, or do behaviors and environment still make a difference?). Research *(20,21)* on adjustment to cancer, as well as to negative life events more generally, suggests that people have a strong tendency to attribute their misfortune to factors within their control. Such beliefs may provide hope that one can reduce the chances of recurrence through one's own behavior, and they are generally associated with good psychologic adjustment, as long as one blames controllable aspects of the situation, such as one's behavior, instead of less mutable aspects, such as one's character or personality. A pressing need exists to communicate advances in genetic testing and other aspects of genetic risk in ways that do not lead people to feel hopeless about their increased risk.

Assess Constellations of Risk Perceptions and Related Health Beliefs

The preceding sections discussed responses to cancer risk communications in terms of risk status, individual differences, and beliefs regarding the mutability of one's risk. These factors highlight a need to understand the effect of risk communications on a wide set of beliefs related to risk and precautionary behaviors. It will be especially important to be sure that a risk communication that heightens perceived vulnerability does not simultaneously create unfavorable changes in other health beliefs related to precautionary behavior.

This possibility may be especially important in the case of several widely studied individual differences that seem to be associated with different effects for different health beliefs. For example, although it appears from the studies by Gerrard et al. *(1)* that people high in self-esteem are less likely than those low in self-esteem to acknowledge the riskiness of their behavior, high self-esteem may have benefits in terms of people's beliefs about whether they can successfully enact the recommended risk-reduction behavior. Similarly, although people with low self-esteem may not counterargue their risk as vigorously or as successfully as those with high self-esteem, they may not respond as well to risk communications in terms of other beliefs, such as self-efficacy, that have been shown to be important to enacting and maintaining behavior change. Thus, risk communications to increase perceived vulnerability that do not address other health beliefs may easily backfire, making the high self-esteem respondents more defensive and the low self-esteem respondents less confident.

A similar set of questions—and potential pitfalls—arises in research on optimistic beliefs. A great deal of research has focused on the possibility that optimists will underestimate their risks of negative events and, correspondingly, on the development of interventions to increase their perceived risk. However, optimism appears to be associated with increased, rather than decreased, attention to health-risk information *(22)* and to more constructive responses to illness and other stressful events *(23)*, suggesting that it is beneficial on balance. The corresponding question of whether pessimism may undermine risk-reduction behavior has yet to be fully examined. In several studies reviewed by Scheier and his colleagues *(23)*, people who are pessimistic are more likely to engage in avoidant coping, denial, and disengagement when faced with stressful circumstances. People who are pessimistic may acknowledge their risk to a greater extent (and perhaps even overestimate it, although this has yet to be determined), but they may fail to act to reduce their risk because they do not believe such responses will be successful in reducing their risk. As these examples illustrate, it may be important to consider the relation of any given individual difference to a wide range of relevant beliefs that may affect precautionary behaviors, not just perceived risk, and to develop risk communications and other interventions that address these related health beliefs.

Assess Multiple Influences on Responses to Risk Communications Over Time

Both target articles *(1,2)* listed several methodologic reasons why cross-sectional studies fail to provide convincing or useful information about the relation between risk perceptions and precautionary behavior. To these vital and frequently overlooked points, I would like to add another point that has been virtually unaddressed in the literature. Often we simply do not know what happens in the interim between exposure to a risk communication and some subsequent behavioral outcome that would explain a positive, negative, or null relation between risk perceptions and precautionary behavior. There are reasons to believe that several factors may influence such relations, including how the risk communication was processed at the time of exposure, the success or failure of subsequent efforts to enact risk-reducing behavior, the processing of subsequent risk-related information, and social influences on both responses to risk information and precautionary behavior. A comprehensive review of such factors is beyond the scope of this paper, but I will provide some examples to illustrate how these factors may influence risk perceptions, precautionary behavior, and subsequent processing of risk communications.

An important initial question may be, "Where does the risk information go and why?" That is, what determines whether a risk communication is counterargued or weakened by conflicting information, as opposed to being elaborated and bolstered by additional consistent information? Interesting and potentially important differences in recall for risk-related information can be found over very short periods of time, yet little is known about how such differences are created and maintained. For

example, in a 1-week follow-up to the study of women who were frequent caffeine drinkers *(18)*, those who did not complete the affirmation reported greater numbers of risk-disconfirming facts on a surprise free recall test than those who completed the affirmation. What intervening process(es) maintained such differential recall? It is unlikely that such differences could be explained simply by a single instance of defensive processing or counterarguing at the time of exposure; instead, it is possible that these participants reminded themselves of the reasons caffeine was not harmful to their health with each cup of coffee, thus repeatedly counterarguing the risk communication in the week following exposure. Conversely, how did those who completed the affirmation differ in the thoughts they brought to mind each time a choice about consuming caffeine arose? Understanding how different aspects of the persuasive message may be brought to mind as people make risk-relevant choiees will be essential to understanding how risk communications may influence beliefs and behaviors over time.

A second, closely related factor is understanding how risk perceptions are related to successful or unsuccessful attempts to enact precautionary behaviors. Risk perceptions may indeed prompt initial efforts to undertake precautionary behaviors, but people may experience frustration, embarrassment, distress, or failure in their efforts. These difficulties may influence subsequent risk perceptions. Reeall that the smokers studied by Gerrard et al. *(1)* who failed in their attempts to quit smoking were the ones who subsequently downplayed the risks of cigarette smoking. Risk perceptions and behaviors are likely to be reciprocally related as people encounter risk-relevant choices on a frequent—and even daily—basis. Experience sampling or daily diary studies might be very useful in understanding how people think about risk information following exposure, especially when that information involves daily behavior and choices, such as diet, exercise, and smoking. Identifying the stumbling blocks on the path to precautionary behavior and how such difficulties affect subsequent risk perceptions and processing of risk information is essential.

Third, as these examples suggest, it is important to understand how initial responses to a risk communication may influence the processing of subsequent risk information. Does initial disagreement with a risk eommunication make it more difficult for subsequent messages to be processed veridically? Are such effects compounded with repeated exposure to risk communications? Sueh phenomena may be especially amenable to analysis in terms of theories of responses to persuasion, in that counterarguing makes disagreeing thoughts more accessible and therefore more likely to be evoked by subsequent risk communications. Related topics for future study might include understanding other aspects of how people manage multiple risk messages. People are exposed to repeated messages about diet, exercise, screening, and other health-related topics, not just those from a planned intervention. It will be important to understand how prior message processing affects responses to this new information, as well as people's strategies for dealing with an ever-increasing volume of information about health risks. Do people become discouraged as they realize that they cannot simultaneously address dozens of potential threats to health? Do people choose to focus on one or two risks and ignore the rest? Studies of people's strategies for dealing with such information on a daily basis would likely give insights into such processes. In this vein, the recent NCI focus group report *(14)* on how the public perceives, processes, and interprets risk information is a valuable start. Understanding how these processes may differ as a function of risk status may also reveal some important information about the relation of risk perceptions to multiple message processing over time.

A final faetor that is relatively understudied in understanding responses to risk communications is how the social interactions of people exposed to risk communications bolster or degrade the effects of the message. For cancer and many other illnesses, the beliefs and behaviors of other people have been shown to be highly influential in how people understand their own risk and whether they enact precautionary behavior *(24,25)*. It will be critical to examine how specific risk communications may be supported or discounted by similar others' views and experiences. Such responses may be especially important in understanding responses to risk communications among people at high risk for cancer, as the desire for information from and about other people has been shown to increase sharply as a function of distress or uncertainty. A daily diary study could answer some of these questions by tracking how people share risk-related information with others following exposure to a risk communication and are influenced by their reactions to the information and its recommendations. Such a study would provide valuable information about how social factors, for better or for worse, influence subsequent emotions, thoughts, and behaviors related to the risk. One promising approach to creating social support for risk perceptions and precautionary behaviors is to conduct risk-communication interventions in small peer groups *(16,26)*. In this way, people see that others like them are similarly concerned about their risk. Such approaches may increase support for beliefs and behaviors related to risk reduction, while at the same time decreasing self-presentational concerns that may interfere with precautionary behaviors.

Conclusion

The points raised in the target articles and in this brief discussion of responses to persuasive messages suggest that there are many reasons that people deviate from the ideal sequence in which exposure to risk communications leads to perceived risk and, ultimately, to precautionary behavior. First, there may be a complex series of trade-offs that influence people's responses to risk communications. Factors that increase people's motivation to process a message may interfere with their ability to do so; similarly, factors that increase one's sense of personal vulnerability may also create feelings of pessimism and hopelessness that interfere with risk-reduction behavior. As Rothman and Kiviniemi *(27)* point out, people who acknowledge increased personal vulnerability may also simultaneously change other perceptions of the health threat, such as its seriousness or prevalence. It is unknown how these different factors work together to influence precautionary behavior and subsequent processing of risk-related information.

The recognition of this complex—and often conflicting—set of responses to risk communications suggests several different approaches for future research. One approach would be to subject major risk communication campaigns to careful pilot studies of responses to risk communications by use of multiple measures of information processing, cognitive elaboration, and related health beliefs to understand what people attend to, what thoughts are generated by the message (agreeing? disagreeing?), how the information is evaluated (convincing? trustworthy?),

what people recall following exposure, and how and whether people believe they can modify their risk. It will be essential to conduct these tests in high- and low-risk groups and among people with different risk factors for cancer.

Second, persuasive communications to increase perceived risk of cancer must be sensitive to several factors that, unless handled carefully, could result in extensive counterarguing and, therefore, increased resistance to change. Special care should be taken to present information about risk behaviors that may be self-defining or that have important self-presentational or other interpersonal functions. It is difficult to alter such behaviors without making the recommended risk-reduction behaviors more socially acceptable or normative and without addressing reasons other than health-related ones that people practice a particular health-risk behavior.

A third pressing need is to understand how people manage risk-related information over time and what happens as they try to incorporate this information and the behavioral changes it suggests into their daily lives. Where does the risk information "go," and why do risk perceptions bear an inconsistent—and sometimes counterproductive—relation to precautionary behaviors? These factors may be different for high- and low-risk groups in ways that may inform the design and delivery of risk communications and of other interventions to maintain support for precautionary behaviors. Additionally, individual differences, such as self-esteem and optimism, may play an important role in such responses, but they may have complex relations to subsequent risk perceptions and behavior. Only prospective studies that examine a range of outcome measures can tell us how different responses to risk communication may be related to risk beliefs and precautionary behaviors over time. Such studies are essential in that the target behaviors, such as changes in diet and exercise and adherence to screening recommendations, must be practiced consistently over time to be effective. Daily diary studies may provide valuable information about how people manage risk information in their social interactions and behavioral choices and how such processes affect risk perceptions and precautionary behaviors over time.

Finally, it may be well worthwhile to consider one of the major conclusions of Vernon's *(2)* review. Her suggestion to focus directly on social and structural factors that may influence cancer screening, such as cultural beliefs, socioeconomic status, insurance coverage, and physician recommendations, is an important reminder that precautionary behaviors are not always purely personal or individual but may also depend on a broader context of social and structural support for such behaviors.

References

(1) Gerrard M, Gibbons FX, Reis-Bergan M. The effect of risk communication on risk perceptions: The significance of individual differences. Monogr Natl Cancer Inst 1999;25:94–100.

(2) Vernon SW. Risk perception and risk communication for cancer screening behaviors: a review. Monogr Natl Cancer Inst 1999;25:101–19.

(3) Gerrard M, Gibbons FX, Bushman BJ. Relation between perceived vulnerability to HIV and precautionary sexual behavior. Psychol Bull 1996; 119:390–409.

(4) Weinstein ND. The precaution adoption process. Health Psychol 1988;7: 355–86.

(5) Eagly AH, Chaiken S. The psychology of attitudes. Fort Worth (TX): Harcourt, Brace Jovanovich; 1993.

(6) Petty RE, Cacioppo JT. Communication and persuasion: central and peripheral routes to attitude change. New York (NY): Springer-Verlag; 1986.

(7) Petty RE, Wegener DT. Attitude change: multiple roles for persuasion variables. In: Gilbert DT, Fiske ST, Lindzey G, editors. Handbook of social psychology. 4th ed. Boston (MA): McGraw-Hill; 1998. p. 323–90.

(8) Jepson C, Chaiken S. Chronic issue-specific fear inhibits systematic processing of persuasive communications. J Soc Beh Pers 1990;5:61–84.

(9) Leventhal H. Findings and theory in the study of fear communications. Adv Exp Soc Psychol 1970;5:119–86.

(10) McAlister A, Perry C, Killen J, Slinkard LA, Maccoby N. Pilot study of smoking, alcohol and drug abuse prevention. Am J Public Health 1980;70: 719–21.

(11) Ditto PH, Lopez DF. Motivated skepticism: use of differential decision criteria for preferred and nonpreferred conclusions. J Pers Soc Psychol 1992;63:568–84.

(12) Ditto PH, Scepansky JA, Munro GD, Apanovitch AM, Lockhart LK. Motivated sensitivity to preference-inconsistent information. J Pers Soc Psychol 1998;75:53–69.

(13) Liberman A, Chaiken S. Defensive processing of personally relevant health messages. Pers Soc Psychol Bull 1992;18:669–79.

(14) National Cancer Institute, Office of Cancer Communications. How the public perceives, processes, and interprets risk information: Findings from focus group research with the general public. Bethesda (MD): Natl Cancer Inst; 1998 Jun Report No.: POS-T086.

(15) Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992; 111:455–74.

(16) Bryan AD, Aiken LS, West SG. Increasing condom use: evaluation of a theory-based intervention to prevent sexually transmitted diseases in young women. Health Psychol 1996;15:371–82.

(17) Leary MR, Tchividjian LR, Kraxberger BE. Self-presentation can be hazardous to your health: impression management and health risk. Health Psychol 1994;13:461–70.

(18) Reed MB, Aspinwall LG. Self-affirmation reduces biased processing of health-risk information. Motivation Emotion 1998;22:99–132.

(19) Aspinwall LG. Rethinking the role of positive affect in self-regulation. Motivation Emotion 1998;22:1–32.

(20) Taylor SE. Adjustment to threatening events: a theory of cognitive adaptation. Am Psychol 1983;38:1163–73.

(21) Janoff-Bulman R. Assumptive worlds and the stress of traumatic events: applications of the schema construct. Soc Cogn 1989;7:113–36.

(22) Aspinwall LG, Brunhart SM. Distinguishing optimism from denial: optimistic beliefs predict attention to health threats. Pers Soc Psychol Bull 1996;22:993–1003.

(23) Scheier MF, Carver CS, Bridges MW. Optimism, pessimism, and psychological well-being. In: Chang EC, editor. Optimism and pessimism: theory, research, and practice. Washington (DC): American Psychological Association. In press 1999.

(24) Buunk BP, Gibbons FX, editors. Health, coping, and well-being: perspectives from social comparison theory. Mahwah (NJ): Erlbaum; 1997.

(25) Croyle RT. Appraisal of health threats: cognition, motivation, and social comparison. Cogn Ther Res 1992;16:165–82.

(26) Fisher JD, Fisher WA, Misovich SJ, Kimble DL, Malloy TE. Changing AIDS risk behavior: effects of an intervention emphasizing AIDS risk reduction information, motivation, and behavioral skills in a college population. Health Psychol 1996;15:114–23.

(27) Rothman AJ, Kiviniemi MT. "Treating people with information": An analysis and review of approaches to communicating health risk information. Monogr Natl Cancer Inst 1999;25:44–51.

Note

I thank Meg Gerrard for her helpful comments on an earlier version of this manuscript.

The Effect of Risk Communication on Risk Perceptions: the Significance of Individual Differences

Meg Gerrard, Frederick X. Gibbons, Monica Reis-Bergan

The purpose of this paper is to address the literature on the relation between risk communication and the initiation of health behavior change. More specifically, we examine the evidence that providing risk information is an effective way to change risk perceptions, as well as the more limited evidence that altering risk perceptions influences risk behavior. The paper discusses significant developments in the research on these issues, describes specific studies that represent trends in this research, and discusses methodologic issues important to the development of the field. Although there are relatively few studies that demonstrate causal links between risk communication and behavior change, recent developments in the field point to the importance of tailoring risk communications to the individual characteristics of targets. Such tailoring has taken a variety of forms, including providing individualized feedback concerning risk status or genetic vulnerability and assessing readiness for behavior change. Future intervention efforts should combine individualized risk status feedback with assessment of individual differences in previous behavior and acceptance of personal vulnerability. [Monogr Natl Cancer Inst 1999;25:94–100]

A fundamental assumption underlying efforts to communicate accurate risk information to the public is that individuals will be more capable of making important decisions about precautionary and risk behaviors if they are more knowledgeable about the consequences of those behaviors. The most striking evidence that risk communications can effect a change in behavior is the fact that the prevalence of smoking among adults in the United States dropped by more than ⅓ in the 30 years following the Surgeon General's report on the dangers of smoking *(1)*. More reeently, dissemination of information about sexual transmission of the human immunodeficiency virus (HIV) was followed by a dramatic decrease in unprotected sex among gay men *(2–4)*. Examples such as these offer support for the appealing hypothesis that risk communications can be effective in changing perceptions of risk and, in so doing, can motivate people to initiate effective precautionary behavior.

As intuitive as this motivational hypothesis appears, however, we know relatively little about the causal links between risk communication and behavior change. There is limited evidence that providing risk information is an effective way to change risk perceptions, and even less evidence that altering risk perceptions motivates new health behavior. The purpose of this paper is to examine the literature on these important links. We will not attempt to present an exhaustive review of this literature; rather, we will examine significant developments in the research and describe, in some detail, specific studies that represent these trends. The paper is organized around the following set of questions: 1) Do risk perceptions motivate the initiation of new precautionary behavior or changes in risk behavior? 2) What are the barriers to acknowledging risk? 3) Can risk communications change risk perceptions? 4) Finally, what research needs to be done to increase our understanding of the relation between risk information and behavior change?

Relation Between Risk Perceptions and Precautionary Behavior

Virtually all theories of health behavior include the hypothesis that perceived vulnerability is the major motivational force behind precautionary behavior *(5)*. Furthermore, a number of reviews *(6–8)* of the empirical literature have supported this motivational hypothesis by providing empirical evidence that perceptions of vulnerability are positively related to a wide variety of preventive behaviors in adolescents and adults. Unfortunately, these reviews have generally included both prospective and cross-sectional studies, leading to confusing and often contradictory eonclusions.

Most cross-sectional studies of the relation between perceived vulnerability and precautionary behavior have been based on the assumption that feelings of vulnerability motivate individuals to engage in precautionary behavior. As several researchers *(9–11)* have asserted, however, this is often an erroneous assumption. For example, the motivational hypothesis posits that the perception that one is vulnerable to sunburns will lead to the use of sunscreen and protective clothing. However, once an individual has taken these preeautions, his or her perceived vulnerability should decrease, the result being a negative correlation between precautionary behavior and current estimates of vulnerability. In other words, in many cases, perceptions of vulnerability should be interpreted as reflections of the amount of precautionary or risk behavior a person has already engaged in, rather than as the motivation for that behavior.[1] We did not include cross-sectional studies in this review because interpretation is difficult.

A number of studies have tested the motivational hypothesis by tracking the association between perceptions of vulnerability and subsequent precautionary behavior. In an early review of this literature, Janz and Becker *(7)* concluded that the majority of these studies supported the motivational hypothesis—the more vulnerable people feel, the more likely they are to *subsequently* engage in preeautionary behaviors (e.g., inoculation for swine flu, screening for blood pressure, or use of clinic preventive services). Subsequent reexamination of these studies, however, led to a somewhat different conclusion. More specifically, Montgomery et al. *(12)* suggested that the role of perceived vulnerability is qualitatively different when the behavior in question is simple (e.g., getting an immunization), as opposed to

Affiliation of authors: Department of Psychology, Iowa State University, Ames.

Correspondence to: Meg Gerrard, Ph.D., Department of Psychology, Iowa State University, Ames IA 50011 (e-mail: MGERRARD@IASTATE.EDU).

See "Notes" following "References."

© Oxford University Press

threatening, taxing, or socially complex (e.g., monthly breast self-examination or regular condom use). In fact, when these authors examined the subset of studies by Janz and Becker that investigated negative events with extreme threat, complex precautionary behavior, or both, they found that only 25% of these studies reported a significant relation between perceptions of vulnerability and precautionary behavior.

This conclusion was further supported by the Chicago Multi-Center AIDS Cohort Study (MACS) of HIV preventive behavior *(13)*. This study assessed risk perceptions and risk behavior in 637 gay men at two different points in time, 6 months apart. These authors found that perceived vulnerability at time 1 was related to risk behavior at time 2; however, this effect was explained by the correlation between time 1 perceived risk and time 1 risk behavior. In their words, "the apparent link between perceived risk and longitudinal changes in behavior is actually explained by the covariability of a sense of risk and behavior at [time] 1" *(14)*. Secondary analysis of data from the Los Angeles MACS study *(15)* revealed the same pattern (Aspinwall LG: personal communication, 1994 Feb 22). It would appear then that longitudinal studies of the relation between perceptions of vulnerability and precautionary behavior have also failed to provide convincing support for the motivational hypothesis.

How can this be the case? Is it possible that the rapid and profound reductions in smoking and unprotected sex that followed widespread dissemination of research documenting the health risks associated with these behaviors were not associated with corresponding increases in vulnerability estimates? One answer to this question has to do with the nature of the research that has examined it. In particular, we suggest that methodologic shortcomings in the extant research are at least partly responsible for the lack of clear findings in prospective examinations of the hypothesis. We will discuss two of these problems.

Although warnings about the inappropriateness of interpreting cross-sectional data as support for the motivational hypothesis have been articulated repeatedly in the literature *(7,9–11)*, two other methodologic issues have received far less attention. The first of these issues is the problem of initial covariation of perceived vulnerability and behavior—the problem noted by Joseph et al. in their analysis of the MACS data *(13,14)*. More than a decade after the observation by Joseph et al., there are few longitudinal studies of the relationship between perceptions of risk and subsequent precautionary behavior. Even fewer studies exist in which the authors control for prior precautionary behavior and, by doing so, test the relation between risk perceptions and change in behavior. Without controlling for time 1 behavior, significant correlations between risk perceptions at time 1 and behavior at time 2 could be interpreted as causal when they are, in fact, the result of initial covariation between perceived risk and behavior at time 1 (as was the case in Joseph et al.). This is a common source of misinterpretation of prospective risk data *(9–11)*. Ideally, because the relation between perceived vulnerability and behavior is reciprocal (*see below*), both constructs should be assessed at time 1 and time 2 to determine whether change in behavior covaries with change in risk perceptions.

The second methodologic issue has to do with the match between the risk question and the participants' experience with the behavior and their stage of behavior change *(9–11)*. If the participant is engaging in a precautionary behavior, such as breast self-examination, and intends to continue this practice, the appropriate risk question is different than if he or she is not practicing the precautionary behavior. More specifically, a simple question about the likelihood of dying of a disease, breast cancer for example, can confound perceptions of risk and intentions to continue self-examinations: "Even if I develop breast cancer, I won't die from it because I intend to be vigilant about early detection." Likewise, different questions are required if the participant is engaging in a risk behavior, trying to quit a risk behavior, or has already quit. For example, an adamant nonsmoker would probably answer the question "What is the likelihood that you will suffer from a smoking-related disease?" with "none" because he or she does not intend to start smoking. To separate the effects of risk perceptions from those of intentions and past behavior, it is often necessary to ask a hypothetical or "conditional" risk question, i.e., one that is contingent on performance of the behavior: "What would be the likelihood that you would suffer from a smoking-related disease if you were to start smoking?" For smokers who are preparing to quit smoking, it is appropriate to ask, "What would be the likelihood that you would suffer from a smoking-related disease if you were to continue smoking?" *and* "What would be the likelihood . . . if you were to stop smoking?" *(16)*. Until there are more studies that incorporate these features and thereby take experience, prior changes in behavior, *and* intentions into account, it is premature to draw definitive conclusions from the existing prospective literature on the motivational hypothesis.

Defensive Responses to Health Communications

Another reason why longitudinal studies about risk perceptions and precautionary behavior have failed to provide convincing support for the motivational hypothesis has to do with the way people respond to health communications. Specifically, many people tend to react to risk information in a defensive manner. This is especially true if the message suggests or implies that their current or past behavior has been unwise or unhealthy.

Risk Awareness

People often think that they are less vulnerable to future negative events than are similar others—others whose level of risk, objectively, is comparable. This optimistic perspective has been termed an "illusion of invulnerability," or simply an optimistic bias *(17,18)*. There is evidence, however, that this optimism is at least tempered by reality. In other words, although people may see themselves as less vulnerable than a peer or the average person, their estimates do reflect a reasonable awareness of their personal risk status *(11)*. For example, adult smokers are aware that their risks of smoking-related disease are greater than are those of nonsmokers *(19)*, and college students' risk perceptions have been shown to correlate with their actual probability of experiencing a variety of health hazards *(20,21)*. Adolescents also understand the relationship between smoking and vulnerability to smoking-related diseases, and they can apply this knowledge to themselves. More specifically, increases in adolescents' smoking over time are accompanied by corresponding increases in personal vulnerability to lung disease *(22)*. In short, people, including young people, may not be accurate in an exact sense when estimating the likelihood that their behavior will lead to a specific negative consequence. They do, however, appear to have an idea of the relative risks associated with their behavior.

Defensive Cognitions

In spite of these reality constraints on perceptions of vulnerability, optimistic biases are quite common, and people engage in a variety of cognitive strategies that protect these biases. For example, early research on reactions to medical diagnoses by Croyle et al. *(23–25)* revealed that people tend to respond to threatening health information with a predictable set of defensive "adjustments" in their thinking, i.e., when learning that they are at risk, people often engage in an active process of distorting and discounting the risk information. More recent research *(22)* has revealed that adolescents exhibit this same kind of defensiveness following increases in their smoking. In one study, smoking and cognitions associated with smoking were tracked over a 3-year period in a sample of 477 adolescents. Results indicated that adolescents who increased their smoking engaged in two types of defensive (cognitive) strategies: (*a*) They *normalized* the behavior by increasing their estimates of the prevalence of smoking among their peers, and (*b*) they *avoided thinking* about the potential negative eonsequences of smoking. Likewise, members of smoking cessation groups who relapse demonstrate a similar kind of defensiveness by decreasing their risk estimates to the health risks associated with smoking *(16,26)*. Furthermore, this defensiveness appears to be associated with subsequent increases in smoking *(22)*.

Thus, people are not passive, unbiased recipients of information about their health. Although most people do not engage in delusional thinking or complete denial of the potential consequences of their behavior, they do possess an impressive array of self-protective strategies that can act as buffers against the unwanted implications posed by risk information. In other words, first-order denial of risk is uncommon, but avoiding thinking about the risk and minimizing its significance appear to be common reactions to risk communications. Furthermore, people with certain characteristics are more likely to engage in these strategies than others. In fact, several personality constructs have been shown to predict reactions to risk communications. For example, individuals characterized by a general tendency to seek threatening information rather than to avoid it (high monitors) respond to risk communications with excessive anxiety that results in avoidance coping and may ultimately be maladaptive in terms of behavior change [*(27,28); see* Salovey et al. *(29)* for a review of personality factors related to reactions to risk information].

A series of studies *(30–32)* has explored the relation between defensiveness and awareness that one's precautionary behavior is less than optimal. This research has demonstrated that people engage in defensive strategies primarily when their risk behavior is made salient. Furthermore, people with high self-esteem find the implication that they are engaging in inappropriate or unwise behavior to be more aversive than do those with low self-esteem. Thus, they are more likely to become defensive in response to risk messages that make their shortcomings salient. In contrast, people with low self-esteem are not as surprised to learn that their health behavior is less than optimal and respond to such information with less defensiveness *(31–33)*.

Self-Esteem and Defensiveness

The important question here is whether these individual differences in defensiveness affect health; in other words, does it interfere with salutary behavior? To answer this question, we examined changes in risk estimates among adult participants in smoking cessation groups conducted by the American Lung Association *(16)*. Perceptions of vulnerability to smoking-related diseases, commitment to quitting, and self-concept were assessed before, during, and after participation in the clinics. The measure of vulnerability used in this study was conditional—"If you were to continue smoking [start smoking again], what would be the likelihood that you would develop . . ." (followed by a series of smoking-related diseases, each answered on a 0%–100% scale). No differenees were seen in the risk pereeptions of high and low self-esteem participants at the beginning of the study, and no changes in the risk perceptions of the abstainers or of the low esteem relapsers between the quit date and the 6-month follow-up. Consistent with our other studies of self-esteem and risk information, however, many of the high esteem relapsers adjusted their perceptions of vulnerability after they resumed smoking—they (significantly) decreased their risk estimates between the quit date and the follow-up. In addition, there was a significant correlation between these changes in risk estimates and commitment to future smoking cessation attempts. More speeifically, the relapsers who demonstrated the most defensiveness (i.e., decreases in risk estimates) reported lower levels of commitment to future attempts to quit. Thus, it appears that the defensive cognitive strategies exhibited by high esteem individuals may soothe their egos but may also inhibit adaptive behavior change.

To summarize, the evidence on defensive reactions to risk communications is mixed. On the one hand, it appears that people are *relatively* accurate regarding risk and tend not to delude themselves about either the probability of negative consequences or the potential severity of those consequences. On the other hand, some people, namely those with high self-esteem who have engaged in risky behavior, are particularly facile at minimizing their vulnerability when their risk behavior is made salient to them, and this process may have adverse behavioral consequences. Thus, it appears that, for risk information to have maximum effect on risk perceptions or risk behavior, it must be presented in a manner that reduces defensiveness and reactance.

So far, then, we have come to somewhat disheartening conclusions regarding two of the questions we posed at the beginning of the paper. First, there is insufficient empirical evidence from longitudinal studies to support the hypothesis that perceptions of vulnerability motivate precautionary behavior. Second, it would appear that a percentage of the population is capable of at least minimizing (if not ignoring) risk information. To address the more general question about whether risk information can have an effect on risk perceptions, we turn to experimental studies of the effects of generic (nontailored) risk communications.

Can Generic (nontailored) Risk Communications Alter Risk Perceptions?

A number of studies have addressed this question by examining the effects of providing people with nonspecific information about health risks with which they are relatively *unfamiliar*. For example, Weinstein and his colleagues *(34,35)* distributed a brochure about the risks associated with radon exposure in communities that had significant radon problems but had not been targeted for media coverage of the problem. They demonstrated that informing homeowners that there was a "substantial probability" of finding high radon levels in homes in their communities significantly increased the homeowners' perceptions of the likelihood that their own homes had unhealthy radon levels.

Mahler et al. *(36)* provided a demonstration of the effect of relatively *familiar* risk information, i.e., information about the risk of skin cancer associated with exposure to the sun, on perceptions of vulnerability. In this study, participants were shown one of two 10-minute slide shows depicting the proper use of sunscreen. One of the slide shows emphasized the use of sunscreen to prevent skin cancer, and the other focused on the use of sunscreen to prevent age spots and wrinkling. The participants who watched either of these slide shows had higher postintervention risk estimates than did those in a no-information control group.

A different strategy has been used to change risk estimates among young women who are engaging in another relatively familiar risk behavior—unprotected sex. In these studies *(30–32)*, women who were engaging in unprotected sexual intercourse were asked to think about the number of times they had failed to use appropriate protection. This review of previous behavior, which was intended to serve as a reminder that that behavior had been risky, did, in fact, increase the women's estimates of the likelihood that they would contract a sexually transmitted disease or have an unplanned pregnancy. Thus, simply reminding people that their behavior has been risky can increase risk estimates. It should be noted, however, that, in two of these three studies, self-esteem moderated the effect of review on risk estimates, such that the review had a greater effect on the risk estimates of low than high esteem women.

In summary, there is evidence that risk communications that are not tailored to individual differences in risk factors can affect risk perceptions. Furthermore, it appears that these communications can be in the form of information about new risks or reminders of risks about which people are already knowledgeable. The research on individual differences in defensive reactions to such information (e.g., differences in monitoring and blunting and in self-esteem), however, indicates that these nontailored communications are more effective for some people than for others. Concern about such individual differences in reactions to risk information has resulted in recent research exploring the efficacy of risk information that is tailored to the individual. These studies have generally focused on matching the message to either recipients' readiness for behavior change or their risk status.

Matching Communications to Readiness for Behavior Change

Theoretical models that describe health behavior change in terms of discrete stages rather than continua have a long history in health psychology. For example, the Transtheoretical Model of Behavior Change suggests that people go through five distinct stages in the process of quitting a risk behavior or adopting a precautionary behavior. The model suggests that the kinds of interventions that are successful in changing behavior vary from stage to stage *(37)*. Similarly, the Precaution Adoption Process Model outlines different kinds of information that are useful in moving people from one stage of precaution adoption to the next *(38)*. Until recently, however, little research had been directed toward investigating the efficacy of tailoring risk communications to the recipient's stage.

An example of the potential of this approach can be found in an experimental test by Weinstein et al. *(39)* of the effectiveness of risk communications about radon exposure. The study focused on two (stage) transitions relevant to home radon testing: 1) the transition from being undecided about testing one's home (stage 3 in the Precaution Adoption Process Model) to deciding to test (stage 5) and 2) the transition from deciding to test to actually ordering a test (stage 6 in the model). The hypothesis was that risk communications designed solely to increase risk estimates would be effective in getting undecided people to decide to test their homes but were not what was needed to move people who had already decided to buy a test to the next stage, i.e., they would not increase the likelihood that people at stage 5 would move to stage 6. Likewise, it was hypothesized that a message that stressed the ease of testing would have no effect on the attitudes, beliefs, and behaviors of those who had not yet decided to test (movement from stage 3 to stage 5) but would facilitate action among those who had already decided to test (movement from stage 5 to stage 6).

The results were as expected: Risk information was very influential in getting people to decide to buy radon home test kits, but it had no effect on people who had already decided to purchase the kits. Similarly, information on how to conduct the test was successful in getting people to buy and use the kits but only if they had previously decided to conduct the test. Thus, this simple experiment demonstrated that risk information has different effects at different stages of precaution adoption and, therefore, lends support to the utility of stage models in developing tailored risk communications.

Use of Biostatistical Models to Tailor Risk Information

A different kind of tailoring has been designed to take advantage of increasingly accurate biostatistical models of risk projection. These studies typically provide participants with feedback from an individualized health risk appraisal that is designed to estimate their morbidity and mortality risk for various diseases on the basis of their medical history, health habits, and diagnostic assessments, such as blood pressure and cholesterol levels. For example, Avis et al. *(40)* employed this approach to provide individuals with information regarding their risk of heart attack and to determine whether such information could alter optimistic bias. In this study, the participants were asked to rate their own likelihood of having a heart attack or stroke within the next 10 years compared with others of their age and sex. After answering this question, participants completed a health risk appraisal that included weight, cholesterol level, blood pressure, and self-reports of family history of heart disease, smoking behavior, physical activity, and recent life stress. The results of this study were disappointing in that the vast majority of participants (>70%) did not change their risk perceptions during the 7–12 weeks between the health risk appraisal and the follow-up assessment. Although the above-average-risk participants were more likely to decrease their optimistic bias than were the average and below-average risk participants (21% responded to the information by decreasing their optimistic bias), 12% of those who received high-risk feedback actually increased their optimistic bias.

In a similar but more comprehensive study of the use of biostatistical data to alter perceptions of vulnerability, Kreuter and Strecher *(41)* asked adult patients in a primary care setting to report their likelihood of having a heart attack, stroke, or fatal motor vehicle crash or of getting cancer "compared to that of others your age and gender" and then to complete a computer-administered health risk appraisal. The appraisal elicited self-

reports of age, height, weight, body frame, diabetes, blood pressure, cholesterol level, smoking, and alcohol consumption. In addition, women were asked about their family history of breast cancer, age at menarche, age at parity, and history of hysterectomy, mammograms, and Pap smears. On completion of the appraisal, patients were provided with feedback regarding their actuarial risk of 10-year mortality as a result of heart attack, stroke, cancer, and automobile accidents. Six months later, the patients responded to the question about their vulnerability again to assess whether the feedback had an effect on optimistic and pessimistic biases. The results of this study were mixed—risk appraisal feedback reduced optimistic bias for stroke and reduced pessimistic bias for cancer. It did not, however, affect risk estimates for heart attack or motor vehicle crashes or reduce optimistic bias for cancer or pessimistic bias for stroke.[2]

Although these two studies *(40,41)* failed to provide clear evidence that health risk appraisals can affect risk perceptions, both of them employed limited measures of perceived vulnerability. More specifically, both studies employed a measure of comparative risk that assessed participants' perceptions that they were more or less vulnerable than the average person. Thus, these studies were designed to measure the effect of risk appraisals on optimistic bias and pessimistic bias rather than risk estimates, and these biases have been shown to be exceptionally resistant to change *(42)*.

Use of Genetic Biomarkers in Tailoring Risk Communications

The current generation of research on the effect of risk communication is evolving in conjunction with the rapid development of our ability to assess risk accurately with various physiologic measures and biomarkers. A provocative example of this approach is provided by the examination by Lerman et al. *(43)* of the effectiveness of individualized genetic biomarker feedback in changing risk estimates and behavior in adult smokers. This study incorporated biomarker feedback about exposure and genetic susceptibility into a quit smoking intervention. One group of smokers (consultation treatment) received a standardized 60-minute individual smoking cessation consultation that included development of a quitting plan, a menu of quitting strategies, and advice on gaining support from friends and family for the quit effort. The second group (exposure feedback treatment) received the same consultation plus feedback about their personal carbon monoxide level and comparison of their level with typical levels for nonsmokers and ex-smokers. The third group (susceptibility feedback treatment) received the consultation, exposure feedback, and feedback regarding their individual susceptibility to exposure (i.e., CYP2D6 genotyping). For example, smokers whose genotype indicated that they metabolized tobacco carcinogens extensively were told that, because of their genetic make-up, they had a higher risk of developing lung cancer than did other smokers who did not have the same genetic make-up. The results of this study indicated that the susceptibility feedback had a dramatic effect on risk estimates—the exposure plus susceptibility feedback treatment resulted in a sixfold increase in risk estimates relative to the consultation treatment. Thus, this study suggests that biomarkers and other physiologic feedback may be useful tools in providing individualized feedback that can affect risk perceptions and risk behaviors.

Is risk communication an effective way to change risk perceptions? A number of studies have demonstrated that, in spite of optimistic bias and natural defensiveness, providing people with risk information and reminders of their risk behavior can alter their estimates of both familiar and novel risks. These effects have been demonstrated for diverse risks ranging from home radon, unplanned pregnancy and sexually transmitted diseases, and skin damage from the sun, to osteoporosis, colorectal cancer, and the dangers of smoking *(30–32,35,43–45)*. Thus, there is convincing evidence that a variety of approaches to risk communication can be effective in altering risk perceptions. What we do not know for sure as of yet, however, is whether risk information that is tailored to individual differences in risk characteristics and readiness to change is more effective in changing risk perceptions than is nontailored information or whether tailoring can be useful in overcoming defensive responses to risk communications. Although the assumption that tailoring risk information will produce stronger effects is appealing, currently there are insufficient data to either support or refute this assumption.

When Does Changing Risk Perceptions Influence Behavior?

We suggest that asking the simple question of whether changing risk perceptions influences behavior ignores much of what we have learned about the complexity of risk and precautionary behavior over the past 20 years. In particular, it ignores evidence that, because people go through stages of behavior change and precaution adoption, risk information is more relevant at some stages than it is at others. It also ignores the finding that there are individual differences in responses to risk information and, more specifically, that some people respond defensively to this information. Finally, it ignores the finding that complex and multidetermined precautionary behaviors, like breast self-examination and smoking cessation, are less likely to be affected by risk information than are simple behaviors (e.g., getting flu shots). Therefore, we suggest that the appropriate question is not whether risk information is effective in altering risk behavior but rather, "What are the conditions under which risk information is most effective in moving people toward behavior change?"

Consistent with this framing of the question, the Precaution Adoption Process Model suggests that changes in intention will occur only when a change in behavior is perceived to be effective and the problem is perceived to be severe enough to warrant action. In other words, risk information will lead to intention to alter behavior only when it is accompanied by perceptions that the problem is severe and that the behavior change will be efficacious. Efforts to demonstrate this effect under other circumstances are most likely doomed to failure.

Our search of the literature revealed only two studies (one published and the other in preparation) that demonstrate that risk perceptions mediate the effect of risk information on intention to change behavior. First, Fong GT et al. (manuscript in preparation) report that genetic counseling is effective in altering risk estimates of ovarian cancer among women patients at an ovarian cancer clinic and that these changes in risk estimates mediate the relation between actual risk and intentions to have future screening. In this study, both the severity of the cancer and the efficacy of early detection through screening were important components of the counseling session. Similarly, Mahler et al. *(36)* provided participants with information about the probability and severity of skin damage from ultraviolet rays and about the efficacy of

sunscreen use. The information was effective in altering estimates of vulnerability to skin damage, and these perceptions did mediate changes in intention to use sunscreen. Furthermore, these authors were able to follow a subset of their participants and found that their interventions were successful in decreasing the amount of sun exposure in these people, as indicated by melanin skin content 6 weeks later. Although the small number of participants in the follow-up sample did not permit definitive conclusions about the mediation of behavior change by change in risk perceptions, these data are consistent with this interpretation.

The study by Lerman et al. *(43)* described earlier leads to similar conclusions, although these authors did not directly address the role of change in perceived risk as a mediator. They did examine the effects of their intervention by stage of behavior change and concluded that smokers in either biomarker feedback group who were in the preparation stage at the beginning of the study (as defined by the Transtheoretical Model of Behavior Change) were more likely to decrease their cigarette smoking during the 2-month follow-up period than were those in the consultation group.

All three of these studies [*(43,36);* and Fong GT et al.: manuscript in preparation] meet the conditions specified in the Precaution Adoption Process Model in that the interventions included components that addressed the *severity* of the risk and the *efficacy* of behavior change. There is, however, surprisingly little research on *how* perceptions of vulnerability and severity interact with perceptions of the efficacy of behavior change to motivate the adoption of precautionary behavior. Given that most people view cancer as extremely severe, these studies suggest that future research examine the efficacy of risk communications that are personalized with regard to individual differences in personal risk factors, ability, or wiliingness to acknowledge vulnerability and with readiness to change behavior when the disease outcomes are perceived to be severe.

Future Directions for Research

Although there are gaps in our knowledge about the effects of risk communication on risk behavior that need to be filled, there are also exciting developments in the field. The next generation of research can achieve more methodologic precision than the last generation, while taking advantage of new developments in tailoring feedback. More specifically, we propose the following four future research directions:

1) To avoid misinterpretation of results because of covariation of prior perceptions of risk and risk behavior, tests of the effect of cancer risk communications on risk perceptions and precautionary behavior should include (*a*) risk communication interventions compared with no-information control groups, (*b*) preintervention and postintervention assessment of risk perceptions and risk behavior, and (*c*) appropriate statistical analyses to detect the role of risk perceptions as a mediator of changes in behavior, controlling for the covariation of perceived vulnerability and prior behavior.

2) Cancer risk communications do not just provide information about new risks. In many cases, they primarily serve as reminders that one's current or past behavior has been unwise. In these cases, people are not passive recipients of risk information but, instead, can be extremely adept at a variety of cognitive strategies to proteet themselves from information they do not want to hear. Consequently, we need research that furthers understanding of defensive reactions to risk communications. We also need research on interventions that reduce this defensiveness and minimize the psychologic reactance that often occurs when a risk eommunication reminds individuals (especially those individuals with high self-esteem, low monitoring, and high blunting) that their recent behavior has been potentially unhealthy or risky.

3) Much of the literature on risk communication focuses on specific diseases (e.g., breast cancer), or even specific precautionary behaviors (e.g., breast self-examination). This kind of research furthers our knowledge about the effectiveness of various risk communication strategies in altering specific behaviors. It does little, however, to build a body of knowledge that is applicable across diseases and behaviors. If the vast majority of risk researchers focus exclusively on the diseases or risk behaviors that are of interest to them, the result will be a failure to examine questions that are applicable to a variety of diseases and behaviors. For example: Does altering behavior that has short-term versus long-term consequences require different risk communication strategies? Do preventive and screening behaviors require different strategies? We suggest that there is a need for research that compares the effect of various risk communication strategies across different types of diseases and precautionary behaviors and that this research should be designed to explore these differences and advance theory relevant to risk communication across different kinds of behaviors.

4) Finally, we suggest that the exciting new research on tailoring feedback to individual risk characteristics should be expanded to include tailoring on additional individual difference variables. For example, risk appraisals that typically include assessment of health habits, such as exercise and smoking, could be expanded to include assessments of defensiveness about current behaviors (e.g., inflated perceptions of prevalence, minimized perceptions of susceptibility, or misattributions of responsibility), as well as readiness to change specific behaviors. Likewise, interventions that provide feedback on genetic susceptibility to lung cancer could be structured to include measurement of perceived severity and perceived efficacy of smoking cessation to guide counseling strategies for individual participants. Such steps would go a long way toward integrating the theoretical and empirical progress of the past 20 years of research on risk and precautionary behavior, behavior change, risk perception, and risk communication.

References

(1) U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control. Smoking prevalence among adults. Rockville (MD): U.S. Department of Health and Human Services; 1996.

(2) Martin JL. AIDS risk reduction recommendations and sexual behavior patterns among gay men: a multifactorial categorical approach to assessing change. Health Educ Q 1986;13:347–58.

(3) McKusick M, Wiley JA, Coates TJ, Stall R, Saika G, Morin S, et al. Reported changes in the sexual behavior of men at risk for AIDS, San Francisco, 1982–84—the AIDS Behavioral Research Project. Public Health Rep 1985;100:622–9.

(4) Winkelstein W Jr, Samuel M, Padian NS, Wiley JA, Lang W, Anderson R, et al. The San Francisco Men's Health Study: III. Reduction in human immunodeficiency virus transmission among homosexual/bisexual men, 1982–86. Am J Public Health 1987;77:685–9.

(5) Weinstein ND. Testing four competing theories of health-protective behavior. Health Psychol 1993;12:324–33.

(6) Becker MH. The health belief model and personal health behavior. Health Educ Monogr 1974;2:409–19.

(7) Janz NE, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11:1–47.
(8) Harrison JA, Mullen PD, Green LW. A meta-analysis of studies of the Health Belief Model with adults. Health Educ Res 1992; 7:107–16.
(9) Gerrard M, Gibbons FX, Bushman BJ. Relation between perceived vulnerability to HIV and precautionary sexual behavior. Psychol Bull 1996; 119:390–409.
(10) Weinstein ND, Rothman, AJ, Nicolich, M. Using correlations to study relationships between risk perceptions and preventive behavior. Psychol Health 1998;13:479–501.
(11) Weinstein ND, Nicolich MM. Correct and incorrect interpretations of correlations between risk perceptions and risk behaviors. Health Psychol 1993; 12:235–45.
(12) Montgomery SB, Joseph JG, Becker MH, Ostrow DG, Kessler RC, Kirscht JP. The Health Belief Model in understanding compliance with preventive recommendations for AIDS: how useful? AIDS Educ Prev 1989;1:303–23.
(13) Joseph JG, Montgomery SB, Emmons C, Kirscht JP, Kessler RC, Ostrow DG, et al. Perceived risk of AIDS: assessing the behavioral and psychological consequences in a cohort of gay men. J Appl Soc Psychol 1987;17:231–50.
(14) Joseph JG, Montgomery SB, Emmons C, Kessler RC, Ostrow DG, Wortman CB, et al. Magnitude and determinants of behavioral risk reduction: longitudinal analysis of a cohort at risk for AIDS. Psychol Health 1987;1:73–96.
(15) Aspinwall LG, Kemeny ME, Taylor SE, Schneider SG, Dudley JP. Psychosocial predictors of gay men's AIDS risk-reduction behavior. Health Psychol 1991;10:432–44.
(16) Gibbons FX, Eggleston TJ, Benthin AC. Cognitive reactions to smoking relapse: the reciprocal relation between dissonance and self-esteem. J Pers Soc Psychol 1997;72:184–95.
(17) Weinstein ND. Why it won't happen to me: perceptions of risk factors and susceptibility. Health Psychol 1984;3:431–57.
(18) Weinstein ND. Unrealistic optimism about susceptibility to health problems: conclusions from a community-wide sample. J Behav Med 1987;10:481–500.
(19) McCoy SB, Gibbons FX, Reis TJ, Gerrard M, Luus CA, Sufka AV. Perceptions of smoking risk as a function of smoking status. J Behav Med 1992;15:469–88.
(20) Gerrard M, Luus CE. Judgments of vulnerability to pregnancy: the role of risk factors and individual differences. Pers Soc Psychol Bull 1995;21:160–71.
(21) Rothman AJ, Klein W, Weinstein ND. Absolute and relative biases in estimations of personal risk. J Appl Soc Psychol 1996;26:1213–36.
(22) Gerrard M, Gibbons FX, Benthin AC, Hessling RM. A longitudinal study of the reciprocal nature of risk behaviors and cognitions in adolescents: what you do shapes what you think, and vice versa. Health Psychol 1996; 15:344–54.
(23) Croyle RT, Ditto PH. Illness cognition and behavior: an experimental approach. J Behav Med 1990;13:31–52.
(24) Croyle RT, Jemmott JB III. Psychological reactions to risk factor testing. In: Skelton JA, Croyle RT, editors. Mental representation in health and illness. New York (NY): Springer-Verlag; 1991. p. 85–107.
(25) Croyle RT, Sande GN. Denial and confirmatory search: paradoxical consequences of medical diagnosis. J Appl Soc Psychol 1988;18:473–90.
(26) Gibbons FX, McGovern PG, Lando HA. Relapse and risk perception among members of a smoking cessation clinic. Health Psychol 1991;10:42–5.
(27) Miller SM. Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. J Pers Soc Psychol 1987;52: 345–53.
(28) Miller SM, Mangan CE. Interacting effects of information and coping style in adapting to gynecologic stress: should the doctor tell all? J Pers Soc Psychol 1983;45:232–6.
(29) Salovey P, Rothman AJ, Rodin J. Health behavior. In: Gilbert DT, Fiske ST, Lindzey G, editors. The handbook of social psychology. Boston (MA): McGraw-Hill; 1998. p. 633–83.
(30) Gerrard M, Gibbons FX, Warner TD. Effects of reviewing risk-relevant behavior on perceived vulnerability among women marines. Health Psychol 1991;10:173–9.
(31) Smith GE, Gerrard M, Gibbons FX. Self-esteem and the relation between risk behavior and perceptions of vulnerability to unplanned pregnancy in college women. Health Psychol 1997;16:137–46.
(32) Boney-McCoy B, Gibbons FX, Gerrard M. Self-esteem, compensatory self-enhancement, and the consideration of health risk. Pers Soc Psychol Bull 1999;25:954–65.
(33) Eggleston TJ. Altering sexual prototypes: an experimental analogue to a sexual intervention program [dissertation]. Ames (IA): Iowa State University; 1997.
(34) Sandman PM, Weinstein ND. Predictors of home radon testing promotion programs. Health Educ Q 1993; 20:471–87.
(35) Weinstein ND, Sandman PM, Roberts NE. Determinants of self-protective behavior: home radon testing. J Appl Soc Psychol 1990;20:783–801.
(36) Mahler HI, Fitzpatrick B, Parker P, Lapin A. The relative effects of a health-based versus an appearance-based intervention designed to increase sunscreen use. Am J Health Promot 1997;11:426–9.
(37) Prochaska JO, DiClemente CC. Stages and processes of self-change in smoking: toward an integrative model of change. J Consult Clin Psychol 1983;51:390–5.
(38) Weinstein ND. The precaution adoption process. Health Psychol 1988;7: 355–86.
(39) Weinstein ND, Lyon JE, Sandman PM, Cuite CL. Experimental evidence for stages of health behavior change: the precaution adoption process model applied to home radon testing. Health Psychol 1998;17:445–53.
(40) Avis NE, Smith KW, McKinlay JB. Accuracy of perceptions of heart attack risk: what influences perceptions and can they be changed? Am J Public Health 1989;79:1608–12.
(41) Kreuter MW, Strecher VJ. Changing inaccurate perceptions of health risk: results from a randomized trial. Health Psychol 1995;14:56–63.
(42) Weinstein ND, Klein WM. Resistance of personal risk perceptions to debiasing interventions. Health Psychol 1995;14:132–40.
(43) Lerman C, Gold K, Audrain J, Lin TH, Boyd NR, Orleans CT, et al. Incorporating biomarkers of exposure and genetic susceptibility into smoking cessation treatment: effects on smoking-related cognitions, emotions, and behavior change. Health Psychol 1997;16:87–99.
(44) Wurtele SK. Increasing women's calcium intake: the role of health beliefs, intentions, and health value. J Appl Soc Psychol 1988;18:627–39.
(45) Blalock SJ, DeVellis BM, Afifi RA, Sandler RS. Risk perceptions and participation in colorectal cancer screening. Health Psychol 1990;9:792–806.

Notes

[1]It should also be noted that many screening behaviors (e.g., breast cancer screening) should not affect perceptions of future vulnerability.

[2]The advisability of altering perceptions that one is more vulnerable than the average person (i.e., pessimistic bias), and thereby possibly increasing risk behavior, and the question of whether accurate risk perceptions are necessary for behavior change are both important issues, but they are beyond the scope of this paper.

Risk Perception and Risk Communication for Cancer Screening Behaviors: a Review

Sally W. Vernon

This review summarizes and synthesizes research findings on risk perception and risk communication related to cancer screening behaviors. The focus is on cancers for which there is evidence that screening reduces mortality, i.e., cervical, breast, and colorectal cancers. The following questions are addressed: 1) Is perceived risk associated with relevant cancer screening behaviors? 2) What factors are associated with perceived risk? 3) Is the relationship between perceived risk and cancer screening behaviors modified by other factors? 4) Have interventions to change perceived risk been effective in modifying risk perceptions? 5) Are these changes related to subsequent cancer screening behaviors? Methodologic issues are discussed, and future research needs are identified. There was consistent evidence that perceived risk was associated with mammography screening, but there were insufficient data on these associations for cervical or colorectal cancer screening behaviors. There was some evidence that perceived risk mediated the association between other variables and screening behaviors; however, because of the small number of studies, the findings are best viewed as hypothesis generating. Studies of interventions to modify risk perceptions provided some support for the view that they are modifiable, but there was conflicting evidence that these changes were related to subsequent cancer screening. Methodologic studies of how best to measure perceived risk are needed. Because most data on the correlates of perceived risk were cross-sectional, it is difficult to determine whether perceived risk is a cause or an effect in relation to cancer screening. Longitudinal studies that measure perceived risk in defined populations with different cancer screening histories and that include follow-up for screening and repeated measurements of risk perception are needed to clarify this relationship. [Monogr J Natl Cancer Inst 1999;25:101–19]

The purpose of this review is to summarize and synthesize research findings on risk perception and risk communication as they relate to cancer screening behaviors. The focus is on cancers for which there is evidence that screening reduces mortality, i.e., cervical, breast, and colorectal cancers. In the case of screening tests or procedures with established efficacy and effectiveness, the goal of risk communication is to encourage or persuade persons to be tested. For screening procedures in which the risks and benefits are uncertain, e.g., mammography screening for women in their forties or prostate-specific antigen testing, the goal is informed decision making. Risk communication about screening behaviors will take different forms, depending on the strength of the scientific evidence establishing the risks and benefits associated with the tests or procedures in question.

Over the past decade, there have been many efforts by public health professionals to persuade age-appropriate women to have mammograms and Pap tests. Many federally funded research projects have developed, implemented, and evaluated theory-based educational interventions to promote the initiation and maintenance of those behaviors *(1–9)*. Table 1 shows data from the Behavioral Risk Factor Surveillance System (BRFSS) on the prevalence of cervical, breast, and colorectal cancer screening behaviors for 1995 *(10)*. The prevalence of "ever" and "recent" Pap testing and mammography screening is relatively high, indicating that efforts to promote screening for breast and cervical cancers have been reasonably successful overall. Because guidelines for colorectal cancer screening have only recently been recommended *(11)*, the dissemination of this information in the population has yet to occur.

Recent reviews have summarized the literature on interventions to promote breast *(12–15)*, cervical *(14,16)*, and colorectal *(17)* cancer screenings. Therefore, the literature on educational interventions to promote cancer screening behaviors is not a focus of this review. Rather, the focus is on risk perception because, as noted by several authors *(18,19)*, perceived risk is a central construct in a number of theories of health behavior [e.g., the Health Belief Model *(20)*, the Precaution Adoption Model *(21,22)*, the Transactional Model of Stress and Coping *(23,24)*, the Self-regulation Model of Health Behavior *(25,26)*, and the Protection Motivation Theory *(27)*]. Risk perception derives from threat appraisal, which is considered to be a major motivating factor in preventive and protective health behaviors. Threat appraisal is based on beliefs about disease risk and severity *(28)*. As defined by Weinstein and Klein *(29)*, perceived risk is one's belief about the likelihood of personal harm. Because risk perception may be an important motivator of a number of health-related behaviors, it is important to understand both the determinants of risk perception and the patterns of association between perceived risk and specific health-related behaviors to develop effective risk communication messages to encourage the adoption of behaviors that will improve health status.

Perceived risk has been used to explain cancer screening behaviors as well as in interventions to promote cancer screenings. However, the literature on perceived risk as it relates to cancer screening behaviors has not been examined systematically across cancer sites. The following terms have been used synonymously in the literature on cancer screening behaviors and are used synonymously here: perceived risk, risk perception, perceived susceptibility, perceived vulnerability, and subjective risk. Data on other social (e.g., socioeconomic status), cognitive (e.g., perceived barriers), and affective (e.g., worry) constructs are discussed as they relate to the relationship between perceived risk and cancer screening, i.e., as mediating or confounding variables. Specifically, the following questions are addressed: 1) Is perceived risk for various cancers associated with relevant

Correspondence to: Sally W. Vernon, Ph.D., School of Public Health, The University of Texas Health Science Center at Houston, P.O. Box 20186, Houston, TX 77225 (e-mail: svernon@utsph.sph.uth.tmc.edu).

See "Note" following "References."

© Oxford University Press

Table 1. Median percentage reporting cancer screening tests or procedures, Behavioral Risk Factor Surveillance System, 1995*

Test or procedure	Median %
Pap test†	
Ever	93.6
Past 3 y	83.6
Mammogram	
Ever‡	81.8
Past 2 y§	69.2
Clinical breast examination	
Ever‡	89.9
Past 2 y§	73.8
Proctoscopy (ever‖)	
Men	41.9
Women	32.9

*Data are from *(10)*.
†Women with an intact uterine cervix who were ≥18 years old.
‡Women ≥40 years old.
§Women ≥50 years old.
‖Women and men ≥50 years old.

cancer screening behaviors? 2) What factors are assoeiated with perceived risk for cancer? 3) Is the relationship, if any, between perceived risk and cancer screening behaviors modified by other factors? 4) Have interventions to change perceived risk been effective in changing or modifying cancer risk perceptions? and 5) Are these changes, if any, related to subsequent cancer screening behaviors? In addition, methodologic issues related to studying perceived risk in the context of cancer screening are discussed, and future research needs are identified.

Three computerized databases were searched from their inception through December 1998: MEDLINE® (from 1966), CANCERLIT® (from 1983), and PsychINFO® (from 1967). Medical subject headings were used to scan titles, abstracts, and subject headings in all databases with the use of the key words "cancer sereening and risk perception," "perceived risk," "perceived susceptibility," "perceived vulnerability," or "subjective risk." The author reviewed all abstracts identified in the search and obtained articles that appeared relevant for more detailed evaluation. Meeting and dissertation abstracts and articles published in a language other than English were excluded. Reference lists of articles selected for inclusion in the review were examined, as were recent tables of contents of journals in which relevant articles were published.

Is Perceived Risk for Various Cancers Associated With Relevant Cancer Screening Behaviors?

McCaul et al. *(30)* performed a meta-analysis of the relationship between perceived breast cancer risk and mammography screening and found that perceived risk was positively associated with mammography screening in 18 of 19 studies. Most of these studies were of women at average risk for breast cancer. The average effect size was $r = 0.16$, adjusted for sample size and was smaller for prospective ($r = 0.10$) compared with cross-sectional studies ($r = 0.19$). There was no support for the hypothesis that there was a curvilinear relationship between perceived risk and screening, i.e., that high and low perceived risk are negatively associated with screening *(30)*. Worry also was positively associated with mammography screening (average weighted effect size was $r = 0.14$), although there were only six studies and the effect sizes ranged from $r = -0.22$ to 0.45 *(30)*. There were few studies of the association between perceived risk and mammography screening among women at increased risk of breast cancer. Generally, the study populations were self-selected [e.g., *(31–33)*], and the results were inconsistent.

In a review of the literature on colorectal cancer screening adherence, Vernon *(17)* found that two *(34,35)* of eight studies reported a positive association between perceived risk and completion of fecal oecult blood test (FOBT), while six studies *(36–41)* reported no association. Three studies *(35,42,43)* examined this association for sigmoidoscopy, and all found a positive association.

Three studies *(44–46)* performed multivariate analysis of a number of cognitive and attitudinal variables, including perceived risk, and cervical cancer screening. After controlling for other variables, one study *(46)* found a positive association with cervical screening, and two studies *(44,45)* found no association.

At this point, there are not enough data to draw firm conclusions about the pattern or magnitude of the associations between perceived risk and cervical cancer screening or any type of colorectal cancer screening. Although the magnitude of the overall effect size was small, studies have found a consistent and positive association between perceived risk and mammography screening in women at average risk of breast cancer *(30)*.

What Factors Are Associated With Perceived Risk for Cancer?

Twelve studies examined correlates of perceived risk for breast cancer *(18,47–54)*, colorectal cancer *(19,55)*, or "any" type of cancer *(56)*. There were no studies of correlates of perceived risk of cervical cancer. Five studies *(19,47,49,52,53)* were of persons at increased risk on the basis of a family history of cancer. Ten *(19,47–52,54–56)* used a cross-sectional design; two *(18,53)* conducted both baseline and follow-up surveys.

Measures of pereeived risk showed some similarity across studies (Table 2). Six studies *(18,19,47,49,50,55)* asked respondents to compare their risk with a reference group, e.g., other women their age. Other measures included asking persons to rate their perceived lifetime chance of developing a specific cancer or asking respondents to rate their risk over a defined time period. Most response formats were Likert-style with 4- to 6-point rating scales.

In studies that examined the association between perceived risk and objective measures of risk (e.g., number of relatives with cancer), one *(47)* found no association; four *(18,49–51)* found a positive association with some, but not all, indicators of objeetive risk; and two *(49,56)* found inconsistent patterns across subgroups. Three studies *(52–54)* of women at increased risk of breast cancer compared a respondent's subjective risk with an objective risk estimate. Among women at increased risk for breast cancer, two studies *(52,54)* found that over 60% overestimated their breast cancer risk compared with Gail model scores *(57)*, whereas another study *(53)* found that only 8% overestimated their breast cancer risk with the use of a method developed by Carter et al. *(58)* to assign objective medical risk. These marked differences may be because of differences in how subjective risk and objective risk were measured, or they may be because of differences in how women were recruited.

Two studies *(47,49)* of first-degree relatives of breast cancer

Table 2. Factors associated with perceived risk*

Author (reference No.)	Study design	Study population	Dependent variable(s) (reference No.)	Independent variable(s) (reference No.)	Results	Comments (reference No.)
Aiken et al. *(18)*	Cross-sectional and prospective surveys in 1987–1989 and in 1991.	Women's community group in the Phoenix area identified through lists of community organizations and networking; 253 groups were invited to participate in an educational intervention to increase mammography screening and BSE; 44 groups accepted. n = 335; 37–77 y old.	Perceived susceptibility to BC measured with two items: "What do you believe is the chance you will contract BC in your lifetime?" and "How susceptible do you feel you are to BC?" Responses were on a 6-point scale from low to high susceptibility. The mean of the two items measured perceived susceptibility. Risk estimate: "What do you believe are your chances of getting BC compared with other women your age?" Responses were on a 5-point scale from a lot lower to a lot higher. Women were then asked an open-ended question about what factors led them to rate their chances of getting BC as (rating from prior question). Factors were classified as risk increasing or decreasing: personal actions, heredity, physiology, environment, psychology, and chance.	Objective risk measures at follow-up: Age MD found lump (past 4 mo) MD told them of BC symptom (past 4 y) Mother or sister had BC All variables except age scored "yes" or "no." Participation in the intervention was controlled. Compliance with 1989 ACS and NCI guidelines.	r = 0.61 between perceived susceptibility and risk estimate measure. Both were positively correlated at follow-up with all objective risk factors except age. Compared with other women their age: lower = 49%; same = 35%; higher = 16%. Mean risk rating was 2.56 (SD = 0.94) and was significantly below the "equal" risk value of 3.0. In cross-sectional analysis at baseline and at follow-up, perceived susceptibility was modestly associated with mammography compliance at baseline (r = 0.12) and at follow-up (r = 0.06). Mean scores of perceived susceptibility did not change over time (mean = 2.7 and 2.8 for baseline and follow-up). In longitudinal analysis, initial compliance predicted perceived susceptibility at follow-up (r = 0.16) controlling for perceived susceptibility at baseline, but perceived susceptibility at baseline was not associated with compliance at follow-up (r = −0.05).	Women were a subset of those studied by Aiken et al. *(63)*. In 1991, re-interviews were attempted with 556 of the 615 women in the original study. Of the 520 still eligible (alive, no BC, and had not moved out of state), 352 were reinterviewed; 335 of them answered the question on subjective risk. Predominantly white, middle-class women.
Audrain et al. *(47)*	Cross-sectional telephone interview. Year of study not given.	FDRs of BC patients recruited through five major cancer centers. Of 532 FDRs, 438 were contacted; 395 were eligible and interviewed.	"In your opinion, compared with other women who do not have a close relative with BC, what are your chances of getting BC someday?" (lower; about the same; a little higher; much higher). Women who answered "lower" or "about the same" were classified as "unaware" of their increased BC risk. 25% were unaware of their increased BC risk.	Age† Race (black vs. white) Marital status Education No. of FDRs with BC Age at menarche Age at 1st live birth No. of previous breast biopsies Relative's age at diagnosis Relative's stage at diagnosis Time between diagnosis and survey Risk notification (no/yes) Time since last mammogram Alcohol consumption Cigarette smoking (yes/no)	NA† OR = 5.5; CI = 2.5–12.0 NA NA NA NA NA NA NA NA NA OR = 2.2; CI = 1.1–42. NA NA OR = 4.0; CI = 1.9–8.3	Multivariate analysis reported here (ORs and 95% CIs). Because of missing data on cancer stage, only 247 of 395 women were included in the analysis. Rerunning the analysis without stage did not essentially change the results.

(Table continues)

Table 2 (continued). Factors associated with perceived risk*

Author (reference No.)	Study design	Study population	Dependent variable(s) (reference No.)	Independent variable(s) (reference No.)	Results	Comments (reference No.)
Blalock et al. *(19)*	Case-comparison study with telephone interviews conducted in 1984–1986.	Two groups of siblings: high risk and average risk. High risk were siblings of CRC patients hospitalized in 1984–1986 (n = 124). Average risk were siblings (matched to CRC siblings on age, race and sex) of patients hospitalized for surgery for nonmalignant conditions during the same time period. n = 171; 40–75 y old.	Absolute perceived susceptibility: How likely are you to get CRC sometime in your lifetime? Rated on a 5-point scale from very unlikely to very likely. Relative perceived susceptibility: What are your chances of getting CRC compared with others your age? Rated on a 5-point scale from a lot lower to a lot higher. Collapsed into "lower," "same," or "higher."	Risk factor perceptions: Why had they rated their chances of getting CRC as they had: personal actions, heredity, physiology, environment, and psychologic attributes. Factors were classified as risk increasing or decreasing, following Weinstein *(22)*.	Absolute perceived susceptibility mean scores for CRC and non-CRC siblings: 3.1 and 2.8 (NS). Relative perceived susceptibility mean scores for CRC and non-CRC siblings: 2.9 and 2.6 ($P<.01$). For the categorical measure, 29% of CRC siblings believed themselves to be at lower risk, and 44% rated themselves as the same as others their age. This distribution was not reported for non-CRC siblings. Completion of FOBT was regressed on risk status (CRC or non-CRC sibling) and relative perceived susceptibility. The latter was associated with FOBT, and the association between risk status and FOBT completion was reduced, but it was still significant. In another model, absolute perceived susceptibility was not associated with FOBT use.	CRC siblings were informed in a letter before the telephone interview that, as a close relative of someone with CRC, they were "somewhat more likely to get this cancer." Although mean scores differed statistically for relative perceived susceptibility, the mean scores for the CRC siblings were at the average; i.e., they saw themselves as "about the same" as others their age. There was little evidence of an optimistic bias operating differentially among CRC and non-CRC siblings. Sibs in both groups showed an optimistic bias regarding their personal actions related to assessing their relative perceived susceptibility to CRC.
Bowen et al. *(54)*	Cross-sectional survey. Year of the study not given.	African-American women were recruited through religious organizations and media channels. n = 113; 18–74 y old.	Accuracy of BC risk status measured by dividing actual medical risk calculated with the use of the model of Gail et al. *(57)* by perceived risk: "On a scale of 0–100, what do you think your chances of getting BC are, where 0 is no chance of getting BC and 100 means you will definitely get it?" Women were categorized into underestimators, overestimators, or extreme overestimators. Intention to get a mammogram measured on a point–point scale.	Demographics: age, ethnic identity (African-American, Black, Afro-American, other), education, marital status, and income. Psychologic: cancer worry, anxiety, depression, coping, and five categories of mental representations as defined by Levanthal and Cameron *(26)* (labels of BC risk, causes of BC risk, symptoms of BC risk, timeline, and consequences). Attitudes towards doctors.	The average Gail score was 8% (SD = 3), and the average perceived risk was 30% (SD = 29%). 41% underestimated, 23% overestimated, and 36% extremely overestimated their BC risk. In univariate analysis of accuracy of perceived risk, there was no consistent pattern in terms of a dose–response association across the three risk groups. Compared with underestimators, the other two groups had higher scores on measures of depression, anxiety, and coping. Intention to have a mammogram was associated with ethnic identity reported as African-American, cancer worry, and anxiety but not with accuracy of risk perception.	Same study population as that in Bowen et al. *(64)*. Study population was composed of volunteers recruited through community organizations and the media who were at low to moderately increased risk for BC or ovarian cancer (<2 FDRs with BC); 46% had more than a college education.

(Table continues)

Table 2 (continued). Factors associated with perceived risk*

Author (reference No.)	Study design	Study population	Dependent variable(s) (reference No.)	Independent variable(s) (reference No.)	Results	Comments (reference No.)
Champion *(48)*	Cross-sectional, population-based mail survey. Year of study not given.	Women >40 y old who participated in a prospective intervention study to increase mammography screening. Of 2822 eligible women, 1404 agreed during a phone call to participate; 581 returned surveys, and 541 were followed up at 1 y. Of the 541 subjects, 404 were ≥40 y old.	1) Stage of mammography adoption based on the transtheoretical model *(59)* and measured following the method of Rakowski et al. *(61)*. 2) Intention to have a mammogram in the next year. 3) Compliance with ACS guidelines.	Perceived susceptibility to BC was measured with a five-item scale; e.g., "I am likely to get BC in the future." Other items were not reported.	Women in action or maintenance scored higher on perceived susceptibility (mean = 16.7; SD = 4.7) than women in precontemplation (mean = 15.4; SD = 5.3) or contemplation (mean = 15.1; SD = 4.8). Although perceived susceptibility was not statistically associated with self-reported past use, women who met ACS guidelines scored higher on perceived susceptibility (mean scores were 15.6 and 16.3 for nonadherent and adherent women). Women who intended to have a mammogram viewed themselves as more susceptible to BC (mean scores were 16.4 and 14.9 for intend and do not intend).	Only univariate results reported. Other HBM variables also were associated with these outcomes but are not summarized here. This is one of the first attempts to relate HBM constructs to stage of change. Predominantly white, middle-class women. Response rate was 41%.
Daly et al. *(52)*	Cross-sectional. Baseline survey conducted between January 1993 and April 1994.	FDRs ≥35 y old of BC patients treated at Fox Chase or Duke University Cancer Centers. n = 969.	Accuracy of risk perception: categorized as the difference (d) between perceived subjective risk and the Gail score *(57)*: accurate = $-0.10 < d < 0.10$; underestimator = $d < -0.10$; overestimator = $0.10 < d$. Collapsed into overestimators vs. underestimators or accurate estimators.	Chance of getting BC rated on a scale from 0% (definitely will not get it) to 100% (definitely will get it). Objective risk: Gail model score.	Overall, 86% overestimated their risk, 11% were accurate, and 3% underestimated their risk. Overestimators were more likely to be (univariate): whites (88%), other (76%); married (89%); widow (73%); employed (88%), not (82%); and <65 y old (~88%), ≥65 y old (70%). Variables not associated with risk perception accuracy: education; No. of FDRs; age at 1st live birth; age at menarche; No. of breast biopsies. Univariate analysis stratified by age (<50 y old, ≥50 y old). For women <50, only race was significant. For women ≥50 y old, employment status and marital status were significant. In multivariate analysis for older women, marital status and employment status remained significant. No multivariate analysis was done for younger age group.	This report is based on the baseline data used by Lerman et al. *(69, 70)* and Schwartz et al. *(71)*. There is a discrepancy with Lerman et al. *(69)* in the categories for objective risk. 1065 patients were identified between January 1993 and April 1994. 1011 (63%) had ≥1 FDR. 867 gave permission to contact the FDR. 1196 eligible FDRs. 969 (89%) agreed to participate in a baseline survey. 71% reported being within the guidelines for mammography screening.

(Table continues)

Table 2 (continued). Factors associated with perceived risk*

Author (reference No.)	Study design	Study population	Dependent variable(s) (reference No.)	Independent variable(s) (reference No.)	Results	Comments (reference No.)
Fries et al. *(53)*	Cross-sectional and prospective. Year of study not given.	Women >40 y old at increased risk for BC, who were members of Group Health Cooperative of Puget Sound. n = 659.	Perceived risk was measured on a 4-point scale: none, borderline, moderate, or high. Women also were asked to rate their perceived chance of getting BC on a scale from 0% to 50%. Mammography status was ascertained through medical records within 6 mo after a survey measuring perceived risk, intentions and beliefs about mammography, and mammography history.	Objective risk was based on BC risk factors and classified into no increased risk, borderline risk, moderate risk, or high risk *(58)*. SOC based on the transtheoretical model *(59)* and measured following the method of Rakowski et al. *(61)*.	The distribution of perceived risk was none (15%), borderline (31%), moderate (34%), or high (20%). 55% accurately identified, 38% underestimated, and 8% overestimated their risk. In hierarchic regression analysis, higher perceived risk was associated with older age (r^2 = 0.01), objective risk (r^2 = 0.21), prior mammography (r^2 = 0.04), and SOC (r^2 = 0.01). Education and income were NS. In logistic regression analysis, mammography use at 6 mo was associated with higher perceived risk (OR = 1.2), SOC (OR = 23.4, 15.9, and 12.2 for contemplation, action, and maintenance vs. precontemplation, respectively). Objective risk was NS. Women who had accurate perceived risk were more likely to obtain a mammogram than those who had inaccurate perceptions in all categories of objective risk.	A baseline risk factor survey was sent to 946 women at increased risk of BC. The sample was stratified to include an equal number of women in each of three risk categories (high, moderate, or borderline). Personalized recommendation letters were mailed to 823 women who returned the survey. Letters included BC risk status, benefits of early detection, an invitation to obtain a free mammogram, and how and where to obtain it. Within 2 wk, a follow-up survey on perceived risk and other factors was mailed, and 659 women returned it. 95% were white. Only 10% had less than a high school education, and 37% were college graduates or more. 65% obtained a mammogram within 6 mo. *P* values (<.5), but not confidence intervals, were given for the logistic regression analysis.
Helzlsouer et al. *(56)*	Cross-sectional survey in 1991–1992.	Empolyees at the Johns Hopkins Oncology Center. n = 509; 19–66 y.	Risk of developing any cancer in the next 20 y and the next 40 y from 0% to 100%. Analyzed as mean scores and categorically based on tertiles: high, medium, and low. Mean perceived risk in the next 20 y: men (21%) and women (31%). Actual risk of developing cancer in the next 20 y for a 30-y-old ranged from 2.9% to 4.9% on basis of age and sex.	Age† Current smoker Duration of employment (>5 y) Relative with cancer Friend with cancer Relative and friend with cancer Self-rated health (fair or poor)	NA either group† + men and women + men; NA women + women; NA men NA either group + women: NA men + women; NA men No associations between 20-y risk based on tertiles of risk and mammography (ever or past 2 y), Pap test (past 3 y), FOBT, or sigmoidoscopy (past 3 y).	Multivariate results reported here for 20-y risk using mean risk scores. Analyses were stratified by sex. No attempt to relate an individual's rating of his or her perceived risk to his or her actual age- and sex- specific risk. High base rates of breast (>78%) and cervical (>88%) cancer screenings. Response rate was 65%.
Hughes et al. *(49)*	Cross-sectional telephone survey. Year of study not given.	FDRs ≥35 y old BC patients who were eligible for a randomized trial of BC risk counseling. Patients were from Duke and Fox Chase Cancer Centers. n = 375 (224 white, 125 black).	Impact of relative's BC diagnosis on FDRs' perception of their own risk of BC. Relative's diagnosis made me feel my own risk was lower, the same, a little higher, or much higher. Classified as "higher/much higher" vs. "lower/same." 39% of blacks and 18% of whites were classified as lower/same.	Age† Education Relative type (mother or sister) Age at 1st live birth Relative's stage at diagnosis BSE in past 3 mo Notified about their risk Worries affect mood Worries affect functioning Impact of Event Scale *(72)* Involved in relative's care Personal BC concern Worries about relative	NA either group† NA either group NA either group + whites (later age); NA blacks + whites (later stage); NA blacks NA either group NA either group NA either group NA either group NA either group + whites; NA blacks + blacks; NA whites + both groups ("sometime" vs. "high or low" worry)	Multivariate results reported here were stratified by race. No measures of association given, only *P* values (<.05). Matched on education and age. No. of eligible subjects not given.

(Table continues)

Table 2 (continued). Factors associated with perceived risk*

Author (reference No.)	Study design	Study population	Dependent variable(s) (reference No.)	Independent variable(s) (reference No.)	Results	Comments (reference No.)
Lipkus et al. *(50)*	Cross-sectional telephone surveys at baseline and at 3 mo. Year of study not given.	Women ≥50 y old who were members of Kaiser Foundation Health Plan in North Carolina and who were participating in a telephone counseling intervention to increase mammography screening. n = 364.	Subjective risk: chance of getting BC within the next 10 y compared with other women their age: below average (40%), average (43%), or above average (17%). SOC based on the transtheoretical model *(59)* and measured following the method of Rakowski et al. *(61)*.	*Note:* independent variables listed separately for each of two models. For subjective risk: Gail score† No. relatives with BC No. of breast biopsies Age at menarche Age at 1st live birth Self-report of breast problems For SOC: Gail score† No. relatives with BC No. of breast biopsies Age at menarche Age at 1st live birth Self-report of breast problems Subjective risk Attributions of risk were ascertained for perceived risk and were classified following the method of Weinstein *(22)* with the use of the following categories: heredity, physiology, personal actions, psychology, environment, and chance.	Correlation between objective and subjective risk was r = 0.21 Univariate results for subjective risk: OR = 2.7; CI = 1.65–4.56† OR = 5.9; CI = 2.91–12.11 OR = 1.3; CI = 0.86–1.91 OR = 0.85; CI = 0.60–1.23 OR = 0.88; CI = 0.71–1.09 OR = 1.2; CI = 1.25–2.72 2) Univariate/ multivariate results for SOC: OR = 1.8; CI = 1.09–2.86/ OR = 1.4; CI = 0.86–2.34† OR = 1.9; CI = 1.01–3.70/NA OR = 1.4; CI = 0.94–2.03/NA OR = 1.02; CI = 0.72–1.44/NA OR = 1.02; CI = 0.83–1.25/NA OR = 1.6; CI = 1.08–2.22/OR = 1.4; CI = 0.96–1.97 OR = 1.5; CI = 1.13–1.91/OR = 1.4; CI = 1.04–1.79 Additional multivariate models were run with subjective risk and either pro, con, or decisional balance. Subjective risk was associated with SOC when cons or decisional balance, but not pros, were in the model, but the ORs were of similar magnitude in all three models (ORs = 1.34, 1.32, and 1.29).	No multivariate analysis was done on subjective risk. The OR for Gail score is for a 0.10-unit change, e.g., going from a 10% to 20% a risk (95% CIs). Women who answered "don't know" were not included in the analysis for SOC. Subjective risk served as the primary mediator of SOC. Subjective risk and SOC measured at 3-mo interview. All other data collected at baseline. Not clear if any intervention contacts occurred between baseline and the 3-mo interview. Pros and cons measured using "agreement" rather than "importance."
Lipkus et al. *(55)*	Cross-sectional telephone survey. Year of study not given.	Adults ≥50 y old who used a community health center in North Carolina. 1318 were in the original sampling frame. n = 547.	Chance of getting CRC sometime in your life: lower than average (36%), average (21%), higher (4%), or don't know (37%). Also asked about attributions of risk: hereditary, physiology, personal actions, psychology, and environment.	Model 1: Age (≥70 vs. 50–59)† Sex Ever smoked Current smoker Self-rated health Accuracy of beliefs (none/all) Attributions of risk (personal actions, hereditary, psychologic, vs. don't know) Model 2: Age† Sex Ever smoked (yes/no)	Model 1: "don't know" vs. "any": OR = 2.0; CI = 1.20–3.17† NA NA NA NA OR = 0.74; CI = 0.66–0.85 NA for any type of attribution Model 2: "below average" vs. "average/above average": NA† NA OR = 0.49; CI = 0.27–0.88	Two logistic models (ORs and 95% CIs): 1) "don't know" vs. "any" risk estimate and 2) "below average" vs. "average/above average." Over one third stated that they didn't know their risk of CRC, and 14% of those who rated their perceived risk answered "don't know" to the question on attributions. A high percentage of self-reports of CRC screening were not verified in a medical chart audit. Participants were mostly low-income blacks

(Table continues)

Table 2 (continued). Factors associated with perceived risk*

Author (reference No.)	Study design	Study population	Dependent variable(s) (reference No.)	Independent variable(s) (reference No.)	Results	Comments (reference No.)
				Current smoker	NA	who had been recruited for a randomized trial to promote breast and cervical cancer screenings and smoking cessation and had been interviewed previously.
				Self-rated health	NA	
				Accuracy of CRC beliefs	NA	
				Attributions of risk:		
				Personal actions	OR = 5.8; CI = 2.12–14.11	
				Hereditary	OR = 4.9; CI = 2.05–11.87	
				Psychologic	OR = 2.4; CI = 1.15–5.15	
Vernon et al. *(51)*	Cross-sectional self-report survey in 1987.	Women ≥35 y old who were participating in a BC screening program in Texas. n = 36 000.	Chance of ever getting BC: very small, moderate, high, or very high. The overall distribution was not given.	Age†	—†	Compared "high/very high" vs. "small/moderate." Multivariate results given (ORs and 95% CIs). Self-selected group of women who were younger, better educated, more likely to be white, and more likely to come from urban areas compared with the female population in Texas. Response rate was 56%.
				Household income	NA	
				Education	NA	
				Race (blacks/whites)	OR = 1.4; CI = 1.17–1.74	
				Ever smoker (yes/no)	OR = 1.3; CI = 1.16–1.36	
				Breast biopsy (yes/no)	OR = 1.3; CI = 1.12–1.43	
				Benign breast disease (yes/no)	OR = 1.6; CI = 1.45–1.79	
				Family history of BC (yes/no)	OR = 11.3; CI = 10.34–12.35	
				Parity	NA	
				Age at menarche (≥12/<12)	OR = 1.2; CI = 1.06–1.28	
				Age at menopause	NA	
				Prior mammogram (ever/never)	OR = 1.6; CI = 1.48–1.77	
				Time since last breast exam	NA	
				Practice BSE	NA	

*ACS = American Cancer Society; BC = breast cancer; BSE = breast self-examination; CI = 95% confidence interval; CRC = colorectal cancer; Dx = diagnosis; FDR = first-degree relative; FOBT = fecal occult blood test; HBM = Health Belief Model; MD = physician; NA = no association; NCI = National Cancer Institute; NS = not statistically significant; OR = odds ratio; SD = standard deviation; SOC = stages of change.

†The findings corresponding to each of the independent variables is listed directly across in the results column.

patients found that African-American women were less likely than white women to be aware that they might be at increased risk of breast cancer because of a family history (Table 2). In an analysis stratified by race, Hughes et al. *(49)* found different correlates for perceived risk in the two groups. In studies that included cigarette smoking, all *(47,51,56)* but one *(55)* found a positive association with perceived risk.

Very few studies have examined psychologic or psychosocial measures in relation to pereeived risk (Table 2). Three studies *(48,50,53)* found that subjective risk was positively associated with later stages of change based on the transtheoretical model *(59–61)*. Bowen et al. *(54)* examined the associations between a number of psychologic variables and accuracy of risk perception. Compared with women who underestimated their breast cancer risk, women who overestimated or extremely overestimated their risk had higher scores on measures of depression, anxiety, and coping abilities *(54)*.

Four studies *(18,19,50,55)* asked respondents to state why they rated their risk as they did, and responses were categorized as risk-increasing or risk-decreasing with the use of a classification scheme developed by Weinstein *(22)*. Lipkus et al. *(50)* and Aiken et al. *(18)* examined attributions of perceived risk for breast cancer. Both studies found that heredity was the most frequently cited cause, followed by physiology and personal actions. In both studies, hereditary and physiology were frequently mentioned as risk-increasing factors (by women who perceived their risk as above average) and as risk-decreasing factors (by women who perceived their risk as below average). Environment, psychology, and chance were not frequently mentioned in either study. In Aiken et al. *(18)*, personal actions were cited as a risk-decreasing factor by women who perceived their risk as lower than average but were rarely mentioned as risk-increasing factors in either study.

Blalock et al. *(19)* and Lipkus et al. *(55)* examined attributions for perceived risk of eolorectal caneer. Siblings of colorectal cancer patients (high-risk group) and siblings of general surgical patients (low-risk group) were more likely to view their personal actions as decreasing rather than increasing their risk, indicating that an optimistic bias was not operating differentially between the two groups *(19)*. Physiology was mentioned with equal frequency by both groups as a risk-increasing and risk-decreasing factor. High-risk siblings were more likely to mention heredity as a risk-increasing than as a risk-decreasing factor, whereas low-risk siblings mentioned it with about equal (and low) frequency as risk increasing or risk decreasing. In multivariate analysis of heredity as a risk-increasing factor in the high-risk group, race was the only statistically significant predictor; 29% of white high-risk siblings cited heredity as a risk-inereasing factor compared with 6% of African-American high-risk siblings.

In contrast to other studies of attributions *(18,19,50)*, Lipkus et al. *(55)* found that, in a group of older, predominantly African-American clinic users, most persons attributed their risk to psychologic causes; however, consistent with the other studies, very few respondents cited environmental factors. In multivariate analysis, attributions of risk were associated with perceived risk.

Compared with persons who did not know why they evaluated their risk as they did, persons who cited psychologic causes, heredity, or personal actions were more likely to rate their risk as below average (Table 2).

Measures of perceived risk showed some similarity across studies. However, differences in the composition study populations, in the variables measured, and in the analytic approaches taken made it difficult to compare findings. In most studies, perceived risk was modestly associated with objective measures of risk; however, in three studies of women at increased risk of breast cancer, women were found to greatly overestimate *(52,54)* or underestimate *(53)* their objective risk. Very few studies examined psychologic or psychosocial correlates, but consistent patterns were found in the three studies that examined the association between stages of change based on the transtheoretical model and perceived risk *(48,50,53)*. Other correlates were not examined in enough studies to provide a basis for generalization.

Is the Relationship, if Any, Between Perceived Risk and Cancer Screening Behaviors Modified by Other Factors?

Four reports *(28,46,62,63)* evaluated the direct and mediating effects of perceived risk on screening compliance or on outcomes related to compliance, e.g., intention (Table 3). Two reports by Aiken et al. *(62,63)* used the same study population to examine the relationship of four Health Belief Model constructs with past mammography screening *(63)* and with prospective compliance *(62)*. In cross-sectional analysis, they found an interaction between perceived susceptibility and perceived barriers with compliance *(63)*. In longitudinal analysis, Aiken et al. *(62)* tested hypotheses about the direct and mediating effects of Health Belief Model constructs on steps to compliance, a variable composed of actions related to scheduling and completing a mammogram, including obtaining one. They hypothesized that high scores on perceived susceptibility and severity would lead to perceived benefits and that greater benefits and fewer perceived barriers would lead to compliance with mammography screening. Intention to obtain a mammogram was hypothesized to link health beliefs and compliance. These hypotheses were confirmed (Table 3).

In a cross-sectional study designed to examine the relationship between social structure and social cognition, Orbell et al. *(46)* examined the effects of perceived susceptibility and a number of other social, cognitive, and attitudinal variables on cervical cancer screening (Table 3). In path analysis, perceived susceptibility was directly and indirectly associated with screening status (Table 3). In another report that used data from a subset of the same population, i.e., women who were up to date on cervical cancer screening, Orbell *(28)* examined the role of personal moral obligation ("I think I should have a screening test") and other independent variables, including perceived susceptibility, on the likelihood that women expected to have a screening test in the future. Among women who were up to date on screening, perceived susceptibility was not directly or indirectly associated with future expectations about having a Pap test (Table 3).

These studies provide some evidence for the indirect or mediating role of perceived risk in cancer screening behaviors; however, because there are so few studies and because of limitations in the study designs [all but one *(62)* were cross-sectional], the findings are probably best viewed as hypothesis generating.

Have Interventions to Change Perceived Risk Been Effective in Changing or Modifying Cancer Risk Perceptions? Are These Changes, if Any, Related to Subsequent Cancer Screening Behaviors?

There have been few educational interventions explicitly designed to change cancer risk perceptions; however, several interventions used persuasive messages to increase mammography screening and also examined the effect of those messages on risk perceptions or other cognitive factors believed to influence cancer screening decisions. The study populations included community-based participants *(62,64)*, volunteers from work sites *(65)*, patients in general practice settings *(66,67)*, and women at increased cancer risk *(68–71)*. Three studies of women at increased risk were based on the same study population *(69–71)*. All but one study *(67)* targeted breast cancer screening behaviors *(62,64–66,68–71)*. Some studies used a theoretic model of behavior change to communicate risk information *(62,65)*, whereas others *(67–71)* provided feedback about actual or objective risk on the basis of statistical models of risk, such as the Gail model *(57)*. Theories and models of behavior change that were used as a basis for intervention development included the Health Belief Model *(62,66)* and prospect theory *(65)*.

Aiken et al. *(62)* developed an intervention to increase mammography screening on the basis of four constructs from the Health Belief Model (Table 4). Pretest and posttest scores on perceived susceptibility showed that scores on the posttest measure increased in both intervention groups compared with the pretest measure of susceptibility in the control group. Similarly, both intervention conditions showed a significant increase from pretest to posttest scores on perceived susceptibility before and after controlling for demographic factors. Compliance with mammography at 3 and 6 months was similar in the two intervention groups and was modestly higher than that in the control group after controlling for covariates (Table 4). Siero et al. *(66)* also used the Health Belief Model to evaluate the effect of four messages that manipulated perceived susceptibility and perceived severity on knowledge, attitudes, intention, and behavior related to breast self-examination. One month after the intervention, there were no differences among groups on perceived susceptibility or on other Health Belief Model constructs (Table 4). Banks et al. *(65)* developed intervention messages on the basis of prospect theory to increase mammography screening (Table 4). Two groups of women employed by a large northeastern utility company were randomly assigned to view videos at the work site that emphasized either the gains or the benefits associated with getting a mammogram or the losses or the risks associated with not getting a mammogram. At 12-month follow-up, a higher percentage of women who viewed the video emphasizing loss-framed messages had obtained a mammogram compared with women who viewed the video emphasizing gain-framed messages, and the intervention effect remained when other variables were controlled (Table 4). Scores on perceived risk of breast cancer, however, did not differ in the two groups immediately after the intervention.

Lerman and colleagues *(69–71)* compared the effects of an educational intervention on breast cancer risk comprehension

Table 3. Factors that modify the association between cancer risk perceptions and cancer screening behaviors*

Author (reference No.)	Study design	Study population (reference No.)	Dependent variable	Independent variable(s) (reference No.)	Results	Comments (reference No.)
Aiken et al. *(63)*	Cross-sectional survey in 1987–1989.	Women's community groups in the Phoenix area identified through lists of community organizations and networking. 253 groups were invited to participate in an educational intervention to increase mammography screening and 44 accepted. n = 615 women; ≥35 y old.	Self-reported compliance with guidelines based on age appropriate screening according to ACS and NCI at the time the study was conducted.	Perceived susceptibility to BC measured with four items rated on a 6-point scale: chance contract BC, susceptible compared with other diseases, type person who gets BC, feel susceptible to BC. Other independent variables: benefits, barriers, physician recommendation, regular source of medical care, knowledge, FDR with BC, lump, income, and age.	There was an interaction between perceived susceptibility and barriers. When barriers were low, susceptibility was positively associated with compliance. When barriers were high, there was no association between perceived susceptibility and compliance. In hierarchical regression analysis in which all variables except the HBM constructs were entered in step 1, the addition of the HBM constructs to the model led to a 7% gain in prediction, and both benefits and barriers were statistically significant predictors of compliance, but perceived susceptibility and severity were not. In a separate model that included two measures of objective risk, perceived susceptibility was an independent predictor of compliance.	This was the only study to that date that included all of the HBM components and had multiple-indicator measures of the constructs. The hierarchical approach to regression analysis provides a conservative estimate of the association between the HBM components and compliance. Same study population as that in Aiken *(18,62)*.
Aiken et al. *(62)*	Cross-sectional survey in 1987–1989 with 3- and 6-mo telephone follow-up.	Subset of the study population in Aiken et al. *(18,63)* Data on 295 women 35–74 y old were collected at baseline. 221 were reinterviewed at 3 mo; and 168 were reinterviewed at 6 mo.	Self-reported compliance with guidelines based on age-appropriate screening according to ACS and NCI at the time the study was conducted.	Perceived susceptibility to BC measured with four items rated on a 6-point scale: chance contract BC, susceptible compared with other diseases, type person who gets BC, feel susceptible to BC. Other independent variables: benefits, barriers, physician recommendation, regular source of medical care, knowledge, FDR with BC, lump, income, and age.	The association between the intervention and benefits was accounted for by both a direct path (r = 0.23) and an indirect path (r = 0.26) through susceptibility. The indirect path through severity was NS. The association between susceptibility (r = 0.20) and severity (r = 0.13) with intentions was largely accounted for by the significant indirect paths through benefits (r = 0.36). The direct paths from susceptibility and severity to intentions were NS. The association between susceptibility, severity, and benefits and steps to compliance was largely mediated by intentions. None of the direct paths was significant.	This was the only prospective study evaluating the mediating effects of perceived risk.

(Table continues)

Table 3 (continued). Factors that modify the association between cancer risk perceptions and cancer screening behaviors*

Author (reference No.)	Study design	Study population (reference No.)	Dependent variable	Independent variable(s) (reference No.)	Results	Comments (reference No.)
Orbell et al. *(46)*	Case-comparison interviews. Year of study not given.	Women aged 20–64 y were selected at random from 23 general practices in Scotland on the basis of cervical cancer screening status (up-to-date or never screened). n = 307 in each group.	Had a cervical cancer screening test within the past 3 y vs. never had cervical cancer screening as identified through medical records.	Perceived susceptibility was measured by three statements: "I think I am personally at risk of cervical cancer," "I think most women my age should have a smear test," and "I don't think it is necessary for me to have a smear test" (scored on a 5-point Likert scale from strongly agree to strongly disagree). Other independent variables: perceived severity, aversiveness of test procedure, aversiveness of test result, benefit of peace of mind, benefit of cure, barriers, physician cue, age, number of sexual partners, marital status, and social class.	In multivariate analysis, perceived susceptibility was strongly associated with screening status (β = 0.48; SE = 0.11). Path analysis showed that perceived susceptibility had a direct effect on screening status (β = 0.25) and that several factors were mediated by perceived susceptibility, including number of sexual partners (β = 0.37) and social class (β = −0.12).	
Orbell *(28)*	Cross-sectional survey. Year of study not given.	Women were a subset of *(46)*. 276 women age 20–60 y with an up-to-date screening history.	Behavioral expectation: "How likely is it that you will have a test in the future?" Response format not given but appears to have been measured on a 5-point scale.	Perceived risk measured as two separate items: "I am at risk of cervical cancer" and "I am at least risk of cervical cancer than some women." Other independent variables: personal moral obligation, worry about cervical cancer, efficacy of treatment, previous screening experience (embarrassment or pain), prior positive test result, and a risk index based on number of sexual partners and smoking status. Additional variables were the same as those in Orbell et al. *(46)* above.	In multivariate analysis, only age (β = −0.20), personal moral obligation (β = 0.38), and aversiveness of the test procedure (β = −0.15) were associated directly with behavioral expectation to repeat the test, accounting for 22% of the variance. Social class also was positively associated with a belief that one was at lower risk compared with other women (β = 0.15).	The coefficients in the text and on the figure differ. Coefficients in the figure are reported here. The authors suggest that the lack of an association between perceived risk and future expectation could be because women in the study population had recently engaged in a risk-reduction behavior for a preventable cancer that allayed concerns and led to decreased risk perceptions.

*ACS = American Cancer Society; BC = breast cancer; FDR = first-degree relative; HBM = Health Belief Model; NA = no association; NCI = National Cancer Institute; NS = not statistically significant; SE = standard error.

and related outcomes including mammography completion. The intervention group received an individualized probability estimate of the risk of developing breast cancer on the basis of the Gail model *(57)*, whereas the control group received general information about guidelines for preventive health behaviors, including breast cancer screening (Table 4). There was no statistically significant difference in risk comprehension between the groups at 3-month follow-up, and approximately two thirds of women in both groups continued to overestimate their lifetime risk following risk counseling *(69)*. The effectiveness of this intervention in reducing breast cancer-specific distress as measured by Impact of Event Scale intrusion scores *(72)* and general psychologic distress *(73)* also was evaluated *(70)*. For breast cancer-specific distress, there was a statistically significant interaction between treatment group and education (Table 4). Furthermore, perceived risk (measured by perceived lifetime risk and by improved accuracy of subjective risk estimates) was not found to mediate the effect of the intervention on Impact of Event Scale scores among less educated participants. In a report *(71)* on the effect of the intervention on mammography compliance, there also was an interaction between treatment group and education. Among women with less education, the intervention led to decreased mammography use; the intervention had no effect on mammography use among more educated women (Table 4).

Bowen et al. *(64)* evaluated the effects of an educational intervention designed to make women's risk perceptions more congruent with medical risk as assessed by the Gail model *(57)* and to increase breast cancer screening intentions. The interven-

Table 4. Interventions to change cancer risk perceptions and cancer screening behaviors*

Author (reference No.)	Study design (reference No.)	Study population (reference No.)	Intervention description (reference No.)	Intervention effect (reference No.)	Comments (reference No.)
Aiken et al. *(62)*	Randomized controlled trial with constraints on randomization conducted during 1987–1989. Baseline telephone survey with 3- and 6-mo follow-up.	Women's community groups in the Phoenix area identified through lists of community organizations and networking. 253 groups were invited to participate, and 44 accepted. There were 615 women ≥35 y old; 348 were not in compliance with ACS guidelines. Data on 295 women 35–74 y old were collected at baseline. 221 were interviewed at 3 mo, and 168 were reinterviewed at 6 mo.	Groups were assigned to one of three conditions. The educational program (E) contained components that specifically targeted four HBM constructs (perceived susceptibility, perceived severity, benefits, and barriers). To increase perceived susceptibility, information was provided on prevalence rates and risk factors for BC. The educational plus psychological (EP) program included the E program plus a series of compliance exercises drawn from the social psychologic literature. Women in the E and EP groups completed a pretest, viewed one of the programs, and immediately completed a posttest. They were interviewed at 3 mo and 6 mo. Control group (C) participants completed a pretest and were interviewed at 3 mo and 6 mo.	After controlling for covariates, posttest means of both the E and EP groups were significantly higher than the C pretest mean on knowledge, susceptibility, benefits, and intention to obtain a mammogram. After controlling for covariates, both E and EP groups showed a significant increase from pretest to posttest in knowledge, susceptibility, benefits, and intention, and a significant decline in barriers. Compliance at 3 mo and 6 mo was higher in the E and EP groups compared with C, but E and EP did not differ from each other. In logistic regression that controlled for demographics and HBM constructs, the intervention OR (E and EP combined) was 1.5 (CI = 1.04–2.04) at 3 mo and 1.4 (CI = 1.03–1.93) at 6 mo.	Perceived susceptibility was measured with four items using a 6-point Likert-type format; coefficient alpha was 0.93. The intervention did not have an effect on barriers. The program did not remove the barrier of cost. Cost and access cannot be addressed by educational programs. This raises a question about increasing perceived susceptibility without removing these barriers. This was a subset of the study population in Aiken et al. *(18,63)*.
Alexander et al. *(68)*	On group pretest and posttest. Year of the study not given.	59 women 40–70 y old enrolled in the Breast Cancer Prevention Trial at Dartmouth-Hitchock Medical Center whose age-specific hazard rate was equal to or greater than that of an average 60-y-old woman as defined by the Gail model *(57)*. Most participants were white, middle-class women.	Women assessed their perceived risk using the U-Titer questionnaire and then met with an MD who instructed them about their objective risk on the basis of their Gail score. The MD shared the result of each participant's Gail score both visually and verbally and discussed the risk factors in the Gail formula. Immediately after the educational session, the U-Titer was re-administered. 29 women completed the U-Titer again 2–11 mo later.	Before the educational session, none of the women's perceived risks matched their calculated risk. 54/59 overestimated their risk, 38 by a factor of 3 or more. Initial perceived risk was higher for women under 50 y old than for women ≥50 y old. Median absolute difference between the Gail model risk score and perceived risk was 39% before education and 1% after. At follow-up, the median difference was 4%.	The U-Titer questionnaire was developed for utility assessment. It is an interactive questionnaire designed to minimize biases, such as "anchoring," "framing," or avoidance of small numbers. Because women were at increased risk of BC, they may have been self-selected on the basis of high perceived risk and worry about BC.
Banks et al. *(65)*	Randomized controlled trial conducted during 1992–1993. Included preintervention and postintervention measures of BC attitudes and beliefs including perceived risk. Posttest measures were completed immediately after viewing a video.	Women ≥40 y old who were not adhering to national guidelines for mammography and who were employees of a large northeastern utility company. Of 181 eligible women who expressed interest in the program, 133 were randomly assigned.	Women were randomly assigned to one of three groups. Used prospect theory to develop intervention messages. Prospect theory predicts that risk-averse options are preferred in the domain of gains and riskier options are preferred in the domain of losses. Detection behaviors are believed to be a riskier choice than prevention behaviors; thus, loss-framed messages are believed to be more persuasive in promoting detection behaviors than gain-framed messages. Women viewed a video at the work site with the same factual content, but the persuasive messages related to mammography screening were framed either in terms of loss (risks incurred by not getting a mammogram) or gain (the benefits of getting a mammogram).	According to prospect theory, loss-framed messages may increase perceived risk, causing persons to be more willing to perform risky behaviors, but perceived susceptibility scores were similar in groups receiving loss (mean = 3.3; SD = 1.1) or gain (mean = 3.0; SD = 1.3) messages. There was a nonstatistically significant difference in compliance at 6 mo: 34% for gain-framed and 45% for loss-framed messages. At 12 mo, the difference was 52% for gain-framed and 66% for loss-framed messages ($P<.05$, one-tailed test). In logistic regression that controlled for the effects of other variables, the ORs and (CIs) for loss-framed vs. gain-framed messages at 6 mo and 12 mo were 1.7	The measure of perceived risk of BC was a mean score based on two items (seven response categories from not at all likely to very likely): 1) the likelihood that they would develop BC and 2) the likelihood that they would die. Women were volunteers, most of whom were white, well educated, and employed. No. of eligible women in the target population was not reported.

(Table continues)

Table 4. Interventions to change cancer risk perceptions and cancer screening behaviors*

Author (reference No.)	Study design (reference No.)	Study population (reference No.)	Intervention description (reference No.)	Intervention effect (reference No.)	Comments (reference No.)
			Women were interviewed by telephone at 6 mo and 12 mo after the intervention to ascertain mammography status.	(0.80–3.68) and 1.7 (0.82–3.69). None of the attitude or belief variables were associated with compliance.	
Bowen et al. *(64)*	Randomized controlled trial. Year of the study not given.	African-American women aged 18–74 y were recruited through religious organizations and media channels. 172 were eligible and were mailed a baseline survey; 113 returned the baseline survey.	Women were randomly assigned either to a group psychosocial counseling intervention of 4 weekly sessions (n = 49) or to a delayed counseling arm (n = 64). The four major components of the intervention were BC risk assessment and perception, education, stress management and problem solving, and social support. 6 mo after randomization, women were mailed a follow-up survey; 92 returned the follow-up survey.	The treatment group decreased their perceived risk from baseline (mean = 34.4; SD = 29.6) to follow-up (mean = 20.1; SD = 23.4), but the control group did not (mean = 31.3; SD = 28.0 at baseline and mean = 32.6; SD = 24.3 at follow-up). The intervention had no effect on intentions to obtain a mammogram, to perform clinical breast examination, or to perform breast self-examination.	Perceived risk was measured with a single item, rating the chances of ever getting breast cancer from 0% to 100%. Medical risk was assessed using the Gail model *(57)*. On the basis of these two scores, women were classified as underestimators, overestimators, or extreme overestimators. The study population was composed of volunteers. Of the eligible volunteers, 65% participated in the intervention, and 53% completed the final survey.
Kreuter and Strecher *(67)*	Randomized controlled trial conducted in August 1992.	1317 patients 18–75 y old from eight family medicine practices in North Carolina. Patients were approached in the waiting rooms and asked to complete a self-administered survey; 80% agreed. A follow-up survey was mailed after 6 mo; 86% (n = 1131) completed it.	Patients were randomly assigned within each practice to one of three groups: 1) enhanced Health Risk Appraisal (HRA) feedback consisting of personal risk information derived from the *Healthier People* algorithms and individually tailored behavioral change messages that addressed perceived barriers to and benefits from changing risky behaviors; 2) typical HRA feedback consisting of personal risk information derived from the *Healthier People* algorithms; or 3) no feedback. Feedback was mailed to patients 2–4 wk after completion of the baseline survey.	16% had optimistic biases, 36% had accurate perceptions, and 48% had pessimistic biases. The intervention reduced pessimistic bias for perceived cancer risk but did not reduce optimistic bias. Patients who had pessimistic biases at baseline and who received either enhanced or typical HRA feedback that their risk was lower than average were 36% more likely to have decreased perceived risk at follow-up compared with control patients.	Risks of heart disease, stroke, cancer, and motor vehicle crash were assessed. Only cancer results are reported here. Risk of "any" or "all" cancer was assessed. Results may have differed for specific cancer sites. Perceived and actual risk of 10-y mortality was measured. Perceived risk was measured by asking "Compared with others your same age and sex, how would you rate your risk of having cancer within the next 10 years" (much lower, lower, about average, higher, or much higher than average). Actual risk was assessed with the Carter Center HRA *(75)* program. A patient's risk was classified as higher than average (>10% higher than the population average), lower (>10% lower), or average (within 10%). Analyses were conducted for all patients and for the subset that remembered receiving the intervention. The authors report results for the latter.
Lerman et al. *(69)*	Randomized controlled trial. Year of study not given. Women were interviewed by telephone at baseline and at 3 mo.	FDRs ≥35 y old of BC patients treated at Fox Chase or Duke University Cancer Centers. Of 438 women invited, 227 completed the baseline interview, and 200 of those completed the 3-mo interview.	Women were randomly assigned to receive either individualized breast cancer risk counseling (BCRC) or general health education (GHE) to increase risk comprehension. The interventions included an educational session with a nurse educator and print materials.	26% in the BCRC group compared with 17% in the GHE group improved their risk comprehension (P = .10). Similar proportions of women in both groups (about two thirds) extremely overestimated their BC risk both before and after the intervention.	Risk comprehension was measured by asking women to rate their chances of getting BC during their lifetime on a scale from 0 (definitely will not get it) to 100 (definitely will get it). It was categorized as the difference between perceived subjective risk

(Table continues)

Table 4 (continued). Interventions to change cancer risk perceptions and cancer screening behaviors*

Author (reference No.)	Study design (reference No.)	Study population (reference No.)	Intervention description (reference No.)	Intervention effect (reference No.)	Comments (reference No.)
			The central element of BCRC was the provision of individualized probability estimates of developing BC on the basis of the Gail model *(57)*. The uncertainty associated with risk data was emphasized. Both absolute risk and relative risks were presented. GHE consisted of presentation of guidelines for preventive health-related behaviors, including BC screening, based on the U.S. Preventive Services Task Force and *Healthy People 2000*.	There was a statistically significant change in risk comprehension from baseline to 3 mo in the BCRC group but not in the GHE group.	and the Gail score into accurate, underestimator, overestimator, or extreme overestimator. Women were classified as "improved" if their risk comprehension category changed to a category in the direction of accurate or stayed in the accurate category. Women in the treatment group who showed improved risk comprehension had significantly lower baseline levels of BC worry (mean = 3.6) compared with control subjects (mean = 7.5), indicating that BCRC was less effective among women with higher baseline levels of BC worry. Women who agreed to participate in the trial were more educated and scored higher on a measure of anxiety than those who declined.
Lerman et al. *(70)*	Same as Lerman et al. *(69)*	Same as Lerman et al. *(69)*	Same as Lerman et al. *(69)*. The primary outcomes for this study were BC-specific distress measured by the Impact of Event Scale (IES) *(72)* and general mood disturbance measured by the total mood disturbance score on the POMS *(73)* at the 3-mo follow-up.	For IES, there was an interaction between education and treatment group. Women in the treatment group with less education showed a reduction in IES scores compared with less educated women in the control group. There was no effect of the intervention among more educated women. For the POMS, treatment group was not a significant predictor of 3-mo distress. Perceived risk [measured as in *(69)*] was not found to mediate the effect of BCRC on IES scores among the less educated participants.	Women in the GHE group had lower levels of perceived risk for BC at baseline and had relatives that were diagnosed at a younger age compared with women in the BCRC group. The authors suggest that by reducing BC-specific distress among women with less education, it may be possible to increase their adherence to mammography. They also suggest that more complex measures of perceived risk may be needed to elucidate changes in risk perception that result from BCRC.
Schwartz et al. *(71)*	Same as Lerman et al. *(69)*	Same as Lerman et al. *(69)*.	Same as Lerman et al. *(69)*. Women were followed up after 1 y to ascertain mammography status.	The groups were comparable on mammography use at baseline: 75% and 71% for BCRC and GHE, respectively. BCRC was associated with a decrease in mammography use (OR = 0.4; CI = 0.2–0.8). There was a statistically significant interaction between treatment group and education. Among women with less education, BCRC led to decreased use (OR = 0.4; CI = 0.2–0.8). There was no effect of BCRC among more educated women (OR = 1.1; CI = 0.6–2.0).	The authors had previously shown that BCRC increases risk comprehension *(69)* and reduces breast cancer-specific psychologic distress compared with GHE in less educated women *(70)*.The authors suggest that BCRC may result in decreased mammography use among less educated women.

(Table continues)

Table 4 (continued). Interventions to change cancer risk perceptions and cancer screening behaviors*

Author (reference No.)	Study design (reference No.)	Study population (reference No.)	Intervention description (reference No.)	Intervention effect (reference No.)	Comments (reference No.)
Siero et al. *(66)*	2 × 2 factorial design to test four messages vs. a nonrandomized control group. 1 mo after the pamphlet was sent, a letter announced that an interviewer would "approach" them to conduct a public opinion poll about cancer. 538 were approached, and 447 were interviewed.	Five general practitioners in three villages (A, B, and C) in The Netherlands provided addresses for women 35–70 y old.	Messages emphasized susceptibility (high vs. low) and seriousness (high vs. low) of BC. Women in villages A and B were randomly assigned to receive one of four pamphlets. Women in village C did not receive any pamphlet. Pamphlets used a question-and-answer format and were accompanied by a cover letter that emphasized public health education. Susceptibility was manipulated by a personal or impersonal address and by stressing or not stressing the personal chance of developing BC. Seriousness was manipulated by stressing or not stressing the seriousness of BC, in adding or not adding negative evaluations, and in using words with different evaluations.	Immediately after reading the pamphlet, women who read the high seriousness and high susceptibility pamphlet had a higher perception of seriousness of and susceptibility to BC than women who read the other pamphlets. However, the differences between the pamphlets had no influence on women's degree of anxiety or fear of BC. 1 mo after sending the pamphlets, the health messages had no effect on knowledge, attitudes, intention, or self-reported behaviors.	Not clear if the interview was face to face or by telephone.

*ACS = American Cancer Society; BC = breast cancer; CI = 95% confidence interval; FDR = first-degree relative; HBM = Health Belief Model; MD = physician; OR = odds ratio; POMS = Profile of Mood States; SD = standard deviation.

tion decreased perceived risk (as intended) but had no effect on screening intentions (Table 4).

Alexander et al. *(68)* conducted a pretest and posttest evaluation of an educational intervention that provided feedback about a woman's individual risk of developing breast cancer. The U-Titer questionnaire *(74)* measured subjective risk, and Gail model scores *(57)* assessed objective risk (Table 4). The median absolute difference between the Gail model risk score and perceived risk was 39% before the education session and 1% after (Table 4).

Kreuter and Strecher *(67)* conducted a randomized controlled trial in family practice patients to evaluate the effectiveness of providing feedback about risk of cancer (any type), heart disease, stroke, and motor vehicle crash (Table 4). Feedback was based on a comparison of an individual's objective risk on the Carter Center's Health Risk Appraisal *(75)* with perceived risk for each cause of death so that persons could be classified as overestimating (pessimistic bias) or underestimating (optimistic bias) their risk on the basis of an objective criterion. In comparison with actual risk, perceived risk of cancer was characterized by pessimistic bias. The intervention reduced pessimistic bias for perceived cancer risk but did not reduce optimistic bias (Table 4).

Collectively, these findings provide some support for the effectiveness of persuasive educational messages to change risk perceptions. Six *(62,64,66–69)* of seven studies were successful in changing risk perceptions in the hypothesized direction. Two studies *(62,65)* found some support for the effect of the intervention on cancer screening behaviors (at least in the short term). However, three studies *(64,66,71)* found no effect of the intervention on breast cancer screening intentions or on self-reported behavior.

Future Research Needs

At present, we do not know what are the "best" measures of perceived risk. Therefore, we do not have good estimates of the prevalence of perceived risk for different types of cancer or for groups at different levels of risk, e.g., population risk, family history, or genetic risk. We also do not have good estimates of the extent to which persons overestimate or underestimate their risk and whether these patterns vary by cancer site, by different measures of objective risk status, or by the context in whieh risk information is conveyed, e.g., clinical or research settings and media coverage. As Slovic *(76,77)* pointed out, perceptions of risk are determined not only by unidimensional statistics of risk but also by qualitative characteristics of a particular risk. For instance, risk for preventable cancers may be perceived differently from those that are not. Although a number of recent studies *(29,78–83)* examined whether the method of presenting risk estimates affected responses to questions about risk perception, only two *(82,83)* examined risk estimates for cancer. One other study *(84)* examined the effects of numeracy on women's understanding of the benefits of mammography screening with the use of four quantitative formats and found that accuracy was strongly related to numeracy regardless of the format used to present information. Methodologic studies that included a variety of measures of perceived risk and that examined their relationship to measures of objective risk would contribute to our understanding of how best to measure this construct as well as to our knowledge about the prevalence of perceived risk and the extent to which it is underestimated or overestimated. These studies will be more informative if they are conducted in defined populations at different levels of cancer risk.

From the studies reviewed here, it is difficult to determine whether risk perception is a cause or an effect in relation to cancer screening. Aiken et al. *(18)* compared cross-sectional and longitudinal patterns of association between perceived risk and self-reported mammography compliance. In longitudinal analysis, perceived susceptibility at baseline was not associated with mammography compliance at follow-up ($r = -0.05$), but mammography compliance at baseline predicted perceived susceptibility at follow-up ($r = 0.16$; $P<0.05$) controlling for perceived

susceptibility at baseline. Studies are needed that measure perceived risk in defined populations with different cancer screening histories and that include follow-up for screening and repeated measurements of risk perception to clarify this relationship.

Only six studies *(62,64,65,69–71)* assessed the immediate effects of interventions on cognitive or psychologic processes as well as on subsequent screening behavior. At present, there are not enough studies of any one cancer site to draw conclusions about the direct and mediating effects of those processes in relation to screening compliance or to make comparisons across cancer sites. The effects of these processes could differ for persons at different levels of cancer risk or for cancers that may be preventable through early detection of premalignant lesions, such as cervical and colorectal cancers, and for those, such as breast cancer, where early detection confers a survival benefit but does not prevent the disease.

A potentially important factor that was not examined in relation to risk perception or cancer screening in any of the studies reviewed here is perceived behavioral control. In one of the early studies of predictors of compliance with fecal occult blood testing, DeVellis et al. *(85)* found that perceived behavioral control predicted completion of the test in siblings of colorectal cancer patients but not in siblings of non-colorectal cancer patients. Related concepts that were examined in only a few studies reviewed here were coping style and coping skills. Several investigators *(23,25)* have emphasized that, when raising awareness of a health threat, it is important to provide specific actions to reduce the threat. In a recently published study of first-degree relatives of breast cancer patients that applied this line of thought, Schwartz et al. *(86)* evaluated the effectiveness of an intervention based on problem-solving training *(87)* to reduce breast cancer-specific and general psychologic distress compared with a control group who received general health education. There was no overall effect of the intervention on cancer-specific distress as measured by Impact of Event Scale intrusion and avoidance subscale scores *(72)* or on the measure of general distress *(73)*, although in a *post hoc* analysis, Impact of Event Scale scores decreased in women who reported that they regularly practiced problem-solving training compared with women in the control group and with women who did not regularly practice the intervention skills *(86)*.

Fischhoff et al. *(88)* pointed out that, although there is evidence that risk estimates are subject to bias, there is less evidence showing that these biases result in inappropriate risk decisions or supporting the idea that people are waiting for accurate risk estimates so that they can make decisions. In relation to cancer screening decisions, we know very little about the behavioral consequences of overestimating or underestimating one's risk. Overestimation may result in hypervigilance, leading women to engage in excessive screening behaviors *(32)*, or it may have the opposite effect *(31,89)*. At present, we really do not know what the goal of interventions designed to influence risk perception should be. That is, we do not know if increasing the accuracy of risk perception will lead to the behavioral outcomes we want to promote. Intervention development would benefit from longitudinal descriptive data on changes in risk perception over time in relation to measures of psychologic status, cognitive factors, and screening participation. If risk perception is related to worry, anxiety, or psychologic distress, interventions may be needed to address those affective conditions as well.

There are virtually no data on what people want to know about the risks they face. This information will become increasingly important as technology increases our ability to identify healthy persons who will inevitably, e.g., Huntington's disease, or with a high degree of certainty, e.g., BRCA1/BRCA2 carriers, develop a disease. Identifying someone as at risk in the interest of prevention or early detection can have profound negative implications on a person's quality of life *(90,91)*.

A number of disciplines have made important contributions to our understanding of risk perception, including geography, sociology, political science, anthropology, and psychology *(76)*. To fully understand risk perception and to develop effective risk communications, we need to take into consideration the perspectives represented by those disciplines, including the role of individual differences in personality, emotion, cognitions, culture, and social processes *(88)*. The primary focus in the studies reviewed here was on individual differences in perceived risk and on factors that modify its effects. However, attitudes and beliefs do not develop in a vacuum. From one perspective, an individual's choice is largely determined by social structural conditions. Habits, norms, and beliefs vary between different social groups and are patterned by the social structure, particularly the social class structure, producing similar views of the world *(92)*. These patterns of socialization are reflected in beliefs and attitudes toward health and illness and health care. From a social epidemiologic perspective, there is a causal link between behavioral differences, socioeconomic circumstances, and health status *(92,93)*. The reason socioeconomic status (SES) has been so consistently linked with disease is because it embodies resources like knowledge, prestige, money, and power that can be used to avoid risks for disease and death, for instance by adopting health innovations such as cancer screening *(93)*. Link et al. *(93)* used data from the BRFSS to show how the SES distribution of mammography screening and Pap testing can have the unanticipated consequence of becoming a mechanism that links SES to cervical cancer and breast cancer mortality.

Although the social and cultural context in which risk communication messages are delivered influences not only how messages are understood but also whether or not they are acted on, other factors need to be considered as well. Rundall and Wheeler *(94)* showed that the effects of income on preventive services were mediated not only by perceived susceptibility but also by difficulties in access to services. Data from the BRFSS showed that the absence of insurance coverage was a significant barrier to mammography screening in the United States *(95)*. For 1996–1997, self-reported mammography use for women 40 years old or older was 71% and 46% in women with and without insurance coverage, respectively. In a similar vein, data from the five National Cancer Institute (NCI) Breast Cancer Screening Consortium studies *(96)* showed that the prevalence of recent clinical breast examination and of receiving a physician's recommendation for a mammogram was higher in the two study sites where women were recruited from health maintenance organizations compared with other settings; these women also were more likely to be in the action stage of adoption as classified by the transtheoretical model *(97)*.

As indicated by the data in Table 1, the success of efforts to promote cervical cancer screening raises a question about when risk communication is no longer a primary consideration in promoting the adoption of health-related behaviors. Seat belt use

legislation made it unnecessary to continue the largely unsuccessful attempts at persuading the public to use seat belts by informing them of the risk of having a fatal accident. The high prevalence of cervical and breast cancer screenings may be, in part, a result of successful efforts to embed the tests in the medical care system and to provide insurance coverage for the tests. The data on mammography use by insurance status *(95,96)* support the view that the success of risk communication to promote cancer screening may depend on access to medical care and other factors such as cultural beliefs and values *(98,99)*. The task at hand is to identify those factors for subgroups, like Hispanics, in which attempts to promote screening have been less successful *(16)*.

This review has focused on studies of correlates of perceived risk, on the direct and mediating effects of perceived risk in relation to cancer screening behaviors in individuals, and on risk communications targeted at individuals. It has not addressed the issue of risk communication through the mass media. As Slovic *(76)* pointed out a number of years ago, most lay persons acquire their information about hazards from the news media. This observation is no less true for information about disease risks and about the benefits of health-promoting behaviors. An excellent example of the media's depiction of breast cancer risk is provided by Lupton *(100)*, who evaluated the messages in the Australian press about the disease from 1987 to 1990. These descriptions undoubtedly influenced many women's risk perceptions about breast cancer (not necessarily in a positive way) both by the overt content and by the more subtle messages that were conveyed. We can infer from secular trends showing an increase in mammography screening over the past decade that public health professionals and advocacy groups have succeeded in raising awareness about breast cancer and screening, despite the largely null findings from carefully designed, community-based randomized controlled trials to promote screening *(1,3, 5–8)*.

We need simultaneously to refine risk communication messages targeted at defined subgroups in the population and to improve our ability to effectively use mass communication channels to reach a broader audience. The former approach is likely to be more effective in promoting cancer screening for cancers such as cervical and breast cancers in which the prevalence of screening is high, whereas the latter approach is likely to be more effective, at least initially, in promoting cancers such as colorectal cancer in which the prevalence of screening is low.

References

(1) Burack RC, Gimotty PA, George J, Simon MS, Dews P, Moncrease A. The effect of patient and physician reminders on use of screening mammography in a health maintenance organization: results of a randomized controlled trial. Cancer 1996;78:1708–21.

(2) Champion V, Huster G. Effect of interventions on stage of mammography adoption. J Behav Med 1995;18:169–87.

(3) Reynolds KD, West SG, Aiken LS. Increasing the use of mammography: a pilot program. Health Educ Q 1990;17:429–41.

(4) King ES, Rimer BK, Seay J, Balshem A, Engstrom PF. Promoting mammography use through progressive interventions: is it effective? Am J Public Health 1994;84:104–6.

(5) Taplin SH, Anderman C, Grothaus L, Curry S, Montano DE. Using physician correspondence and postcard reminders to promote mammography use. Am J Public Health 1994;84:571–4.

(6) Skinner CS, Strecher VJ, Hospers H. Physicians' recommendations for mammography: do tailored messages make a difference? Am J Public Health 1994;84:43–9.

(7) Urban N, Taplin SH, Taylor VM, Peacock S, Anderson G, Conrad D, et al. Community organization to promote breast cancer screening among women ages 50–75. Prev Med 1995;24:477–84.

(8) Curry SJ, Taplin SH, Anderman C, Barlow WE, McBride CM. A randomized trial of the impact of risk assessment and feedback on participation in mammography screening. Prev Med 1993;22:350–60.

(9) Rakowski W, Ehrich B, Goldstein MG, Rimer BK, Pearlman DN, Clark MA, et al. Increasing mammography among women aged 40–74 by use of a stage-matched, tailored intervention. Prev Med 1998;27:748–56.

(10) Powell-Griner E, Anderson JE, Murphy W. State- and sex-specific prevalence of selected characteristics—Behavioral Risk Factor Surveillance System, 1994 and 1995. MMWR 1997;46:1–29.

(11) Winawer SJ, Fletcher RH, Miller L, Godlee F, Stolar MH, Mulrow CD, et al. Colorectal cancer screening: clinical guidelines and rationale. Gastroenterology 1997;112:594–642.

(12) Curry SJ, Emmons KM. Theoretical models for predicting and improving compliance with breast cancer screening. Ann Behav Med 1994;16: 302–16.

(13) Rimer BK. Mammography use in the U.S.: trends and the impact of interventions. Ann Behav Med 1994;16:317–26.

(14) Meissner HI, Breen N, Coyne C, Legler JM, Green DT, Edwards BK. Breast and cervical cancer screening interventions: an assessment of the literature. Cancer Epidemiol Biomark Prev 1998;7:951–61.

(15) Rimer BK. Understanding the acceptance of mammography by women. Ann Behav Med 1992;14:197–203.

(16) Marcus AC, Crane LA. A review of cervical cancer screening intervention research: implications for public health programs and future research. Prev Med 1998;27:13–31.

(17) Vernon SW. Participation in colorectal cancer screening: a review. J Natl Cancer Inst 1997;89:1406–22.

(18) Aiken LS, Fenaughty AM, West SG, Johnson JJ, Luckett TL. Perceived determinants of risk for breast cancer and the relations among objective risk, perceived risk, and screening behavior over time. Womens Health Res Gender Behav Policy 1995;1:27–50.

(19) Blalock SJ, DeVellis BM, Afifi RA, Sandler RS. Risk perceptions and participation in colorectal cancer screening. Health Psychol 1990;9: 792–806.

(20) Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11:1–47.

(21) Weinstein ND. The precaution adoption process. Health Psychol 1988;7: 355–86.

(22) Weinstein ND. Why it won't happen to me: perceptions of risk factors and susceptibility. Health Psychol 1984;3:431–57.

(23) Lazarus RS. Psychological stress and coping in adaptation and illness. Int J Psychiatry Med 1974;5:321–33.

(24) Monat A, Averill JR, Lazarus RS. Anticipatory stress and coping reactions under various conditions of uncertainty. J Pers Soc Psychol 1972; 24:237–53.

(25) Leventhal H, Zimmerman R, Gutmann M. Compliance: a self-regulation perspective. In: Gentry WD, editor. Handbook of behavioral medicine. New York (NY): Guilford Press; 1984. p. 369–436.

(26) Leventhal H, Cameron L. Behavioral theories and the problem of compliance. Patient Educ Couns 1987;10:117–38.

(27) Rogers RW. Cognitive and psychosocial processes in fear appeals and attitude change: a revised theory of protection motivation. In: Cacioppo BL, Petty LL, Shapiro S, editors. Social psychophysiology: a sourcebook. New York (NY): Guilford Press; 1983. p. 153–76.

(28) Orbell S. Cognition and affect after cervical screening: the role of previous test outcome and personal obligation in future uptake expectations. Soc Sci Med 1996;43:1237–43.

(29) Weinstein ND, Klein WM. Resistance of personal risk perceptions to debiasing interventions. Health Psychol 1995;14:132–40.

(30) McCaul KD, Branstetter AD, Schroeder DM, Glasgow RE. What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. Health Psychol 1996;15:423–29.

(31) Kash KM, Holland JC, Halper MS, Miller DG. Psychological distress and surveillance behaviors of women with a family history of breast cancer. J Natl Cancer Inst 1992;84:24–30.

(32) Lerman C, Kash KM, Stefanek M. Younger women at increased risk for breast cancer: perceived risk, psychological well-being, and surveillance behavior. Monogr Natl Cancer Inst 1994;16:171–6.
(33) Vogel VG, Graves DS, Vernon SW, Lord JA, Winn RJ, Peters GN. Mammographic screening of women with increased risk of breast cancer. Cancer 1990;66:1613–20.
(34) Farrands PA, Hardcastle JD, Chamberlain J, Moss S. Factors affecting compliance with screening for colorectal cancer. Community Med 1984; 6:12–9.
(35) Price JH. Perceptions of colorectal cancer in a socioeconomically disadvantaged population. J Community Health 1993;18:347–62.
(36) Burack RC, Liang J. The early detection of cancer in the primary-care setting: factors associated with the acceptance and completion of recommended procedures. Prev Med 1987;16:739–51.
(37) Halper MS, Winawer S, Brody RS, Andrews M, Roth D, Burton G. Issues of patient compliance. In: Colorectal caneer: prevention, epidemiology, and screening. New York (NY): Raven Press; 1980. p. 299–310.
(38) Hoogewerf PE, Hislop TG, Morrison BJ, Burns SD, Sizto R. Health belief and compliance with screening for fecal occult blood. Soc Sci Med 1990; 30:721–6.
(39) Macrae FA, Hill DJ, St. John DJB, Ambikapathy A, Garner JF. The Ballarat General Practitioner Research Group. Predicting colon cancer screening behavior from health beliefs. Prev Med 1984;13:115–26.
(40) Myers RE, Ross E, Jepson C, Wolf T, Balshem A, Millner L, et al. Modeling adherence to colorectal cancer screening. Prev Med 1994;23: 142–51.
(41) Sandler RS, DeVellis BM, Blalock SJ, Holland KL. Participation of high-risk subjects in colon cancer screening. Cancer 1989;63:2211–5.
(42) Kelly RB, Shank JC. Adherence to screening flexible sigmoidoscopy in asymptomatic patients. Med Care 1992;30:1029–42.
(43) Lewis SF, Jensen NM. Screening sigmoidoscopy: factors associated with utilization. J Gen Intern Med 1996;11:542–4.
(44) Lerman C, Caputo C, Brody D. Factors associated with inadequate cervical cancer screening among lower income primary care patients. J Am Board Fam Prac 1990;3:151–6.
(45) Murray M, McMillan C. Health beliefs, locus of control, emotional control and women's cancer screening behaviour. Br J Clin Psychol 1993; 32:87–100.
(46) Orbell S, Crombie I, Johnston G. Social cognition and social structure in the prediction of cervical screening uptake. Br J Health Psychol 1996;1: 35–50.
(47) Audrain J, Lerman C, Rimer B, Cella D, Steffens R, Gomez-Caminero A. Awareness of heightened breast cancer risk among first-degree relatives of recently diagnosed breast cancer patients. The High Risk Breast Cancer Consortium. Cancer Epidemiol Biomarkers Prev 1995;4: 561–5.
(48) Champion VL. Beliefs about breast cancer and mammography by behavioral stage. Oncol Nurs Forum 1994;21:1009–14.
(49) Hughes C, Lerman C, Lustbader E. Ethnic differences in risk perception among women at increased risk for breast cancer. Breast Cancer Res Treatment 1996;40:25–35.
(50) Lipkus IM, Rimer BK, Strigo TS. Relationships among objective and subjective risk for breast cancer and mammography stages of change. Cancer Epidemiol Biomarkers Prev 1996;5:1005–11.
(51) Vernon SW, Vogel VG, Halabi S, Bondy ML. Factors associated with perceived risk of breast cancer among women attending a screening program. Breast Cancer Res Treatment 1993;28:137–44.
(52) Daly MB, Lerman CL, Ross E, Schwartz MD, Sands CB, Masny A. Gail model breast cancer risk components are poor predictors of risk perception and screening behaviour. Breast Cancer Res Treatment 1996;41: 59–70.
(53) Fries EA, White KS, Bowen DJ, Taplin S, Montano DE. A prospective study of accuracy of risk perceptions for breast cancer and mammography use. Health Psychol. In press 1999.
(54) Bowen DJ, Hickman KM, Powers D. Importance of psychological variables in understanding risk perceptions and breast cancer screening of African American women. Womens Health Res Gender Behav Policy 1997;3:227–42.
(55) Lipkus IM, Rimer BK, Lyna PR, Pradhan AA, Conaway M, Woods-Powell CT. Colorectal screening patterns and perceptions of risk among African-American users of a community health center. J Community Health 1996;21:409–27.
(56) Helzlsouer KJ, Ford DE, Hayward RSA, Midzenski M, Perry H. Perceived risk of cancer and practice of cancer prevention behaviors among employees in an oncology center. Prev Med 1994;23:302–8.
(57) Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989; 81:1879–86.
(58) Carter AP, Thompson RS, Bourdeau RV, Andenes J, Mustin H, Straley H. A clinically effective breast cancer screening program can be cost-effective too. Prev Med 1987;16:19–34.
(59) Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. Health Psychol 1994;13:39–46.
(60) Prochaska JO, DiClemente CC. Stages of change in the modification of problem behaviors. In: Hersen M, Eisler RM, Miller PM, editors. Progress in behavior modification. Sycamore (IL): Sycamore Publishing Company; 1992. p. 184–206.
(61) Rakowski W, Dube CE, Marcus BH, Prochaska JO, Velicer WF, Abrams DB. Assessing elements of women's decisions about mammography. Health Psychol 1992:11:111–8.
(62) Aiken LS, West SG, Woodward CK, Reno RR, Reynolds KD. Increasing screening mammography in asymptomatic women: evaluation of a second-generation, theory-based program. Health Psychol 1994;13:526–38.
(63) Aiken LS, West SG, Woodward CK, Reno RR. Health beliefs and compliance with mammography-screening recommendations in asymptomatic women. Health Psychol 1994;13:122–9.
(64) Bowen DJ, Christensen CL, Powers D, Graves DR, Anderson CAM. Effects of counseling and ethnic identity on perceived risk and cancer worry in African American women. J Clin Psychol Med Settings 1998: 5;365–79.
(65) Banks SM, Salovey P, Greener S, Rothman AJ, Moyer A, Beauvais J, et al. The effects of message framing on mammography utilization. Health Psychol 1995;14:178–84.
(66) Siero S, Kok G, Pruyn J. Effects of public education about breast cancer and breast self-examination. Soc Sci Med 1984;18:881–8.
(67) Kreuter MW, Strecher VJ. Changing inaccurate perceptions of health risk: results from a randomized trial. Health Psychol 1995;14:56–63.
(68) Alexander NE, Ross J, Sumner W, Nease RF Jr, Littenberg B. The effect of an educational intervention on the perceived risk of breast cancer. J Gen Intern Med 1995;11:92–7.
(69) Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995;87:286–92.
(70) Lerman C, Schwartz MD, Miller SM, Daly M, Sands C, Rimer BK. A randomized trial of breast cancer risk counseling: interacting effects of counseling, educational level, and coping style. Health Psychol 1996;15: 75–83.
(71) Schwartz M, Rimer B, Daly M, Sands C, Lerman C. A randomized trial of breast cancer risk counseling: the impact upon self-reported mammography utilization. Am J Public Health 1999;89:924–6.
(72) Horowitz M, Wilner N, Alvarez W. Impact of event scale: a measure of subjective stress. Psychosom Med 1979;41:209–18.
(73) McNair D, Lorr M, Droppelman L. Profile of Mood States [data file]. San Diego (CA): Educational and Industrial Testing Service; 1971.
(74) Sumner W, Nease RFJ, Littenberg B. U-Titer: a utility assessment tool. In: Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care. Washington (DC): McGraw Hill; 1991. p. 701–5.
(75) Amler R, Moriarty D, Hutchins E, editors. Healthier People [guides and documentation]. Decatur (GA): Carter Center of Emory University Health Risk Appraisal Program; 1989.
(76) Slovic P. Perception of risk. Science 1987;236:280–5.
(77) Slovic P, Fischhoff B, Lichtenstein S. Why study risk perception? Risk Analysis 1982;2:83–93.
(78) Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM. The framing effect of relative and absolute risk. J Gen Intern Med 1993;8: 543–8.
(79) Sandman PM, Weinstein ND, Miller P. High risk or low: how location on a "risk ladder" affects perceived risk. Risk Analysis 1994;14:35–45.

(80) Weinstein ND, Diefenbach MA. Percentage and verbal category measures of risk likelihood. Health Educ Res 1997;12:139–41.
(81) Diefenbach MA, Weinstein ND, O'Reilly J. Scales for assessing perceptions of health hazard susceptibility. Health Educ Res 1993;8:181–92.
(82) Rothman AJ, Klein WM, Weinstein ND. Absolute and relative biases in estimations of personal risk. J Appl Soc Psychol 1996;26:1213–36.
(83) Hallowell N, Statham H, Murton F, Green J, Richards M. "Talking about chance": the presentation of risk information during genetic counseling for breast and ovarian cancer. J Genet Counseling 1997;6:269–86.
(84) Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med 1997;127:966–72.
(85) DeVellis BM, Blalock SJ, Sandler RS. Predicting participation in cancer screening: the role of perceived behavioral control. J Appl Soc Psychol 1990;20:639–60.
(86) Schwartz MD, Lerman C, Audrain J, Cella D, Rimer B, Garber J, et al. The impact of a brief problem-solving training intervention for relatives of recently diagnosed breast cancer patients. Ann Behav Med 1998;20:7–12.
(87) D'Zurilla TJ. Problem-solving therapy: a social competence approach to clinical intervention. New York (NY): Springer; 1988.
(88) Fischhoff B, Bostrom A, Quadrel MJ. Risk perception and communication. Ann Rev Public Health 1993;14:183–203.
(89) Lerman C, Daly M, Sands C, Balshem A, Lustbader E, Heggan T, et al. Mammography adherence and psychological distress among women at risk for breast cancer. J Natl Cancer Inst 1993;85:1074–80.
(90) Sachs L. Is there a pathology of prevention? The implications of visualizing the invisible in screening programs. Cult Med Psychiatry 1995;19: 503–25.
(91) Sachs L. Risk as diagnosis: implications for the quality of life. In: Levy et al., editor. Cancer, AIDS, and quality of Life. New York (NY): Plenum Press; 1997. p. 107–13.
(92) Lindbladh E, Lyttkens CH, Hanson BS, Ostergren P, Isacsson SO, Lindgren B. An economic and sociological interpretation of social differences in health-related behavior: an encounter as a guide to social epidemiology. Soc Sci Med 1996;43:1817–27.
(93) Link BG, Northridge ME, Phelan JC, Ganz ML. Social epidemiology and the fundamental cause concept: on the structuring of effective cancer screens by socioeconomic status. The Milbank Q 1998;76:375–402.
(94) Rundall TG, Wheeler JRC. The effect of income on use of preventive care: an evaluation of alternative explanations. J Health Soc Behav 1979; 20:397–406.
(95) Centers for Disease Control and Prevention. Self-reported use of mammography and insurance status among women aged ≥40 years—United States, 1991–1992 and 1996–1997. MMWR 1998;47:825–30.
(96) Stoddard AM, Rimer BK, Lane D, Fox SA, Lipkus IM, Luckmann R, et al. Underusers of mammogram screening: stage of adoption in five U.S. subpopulations. The NCI Breast Cancer Screening Consortium. Prev Med 1998;27:478–87.
(97) Rakowski W, Ehrich B, Dube CE, Pearlman DN, Goldstein MG, Peterson KK, et al. Screening mammography and constructs from the Transtheoretical Model: associations using two definitions of the stages-of-adoption. Ann Behav Med 1996;18:91–100.
(98) Rajaram SS, Rashidi A. Minority women and breast cancer screening: the role of cultural explanatory models. Prev Med 1998;27:757–64.
(99) Fiscella K, Franks P, Clancy CM. Skepticism toward medical care and health care utilization. Med Care 1998;36:180–9.
(100) Lupton D. Femininity, responsibility, and the technological imperative: discourses on breast cancer in the Australian press. Int J Health Serv 1994;24:73–89.

Note

I thank Colette Miesse for her contributions to many aspects of the work on this paper, including bibliographic searches and assistance with compiling data from the literature, and Brenda Brown for her assistance in preparing the final version of the manuscript.

Persuasion for the Purpose of Cancer Risk Reduction: a Discussion

Peter Salovey, Tamera R. Schneider, Anne Marie Apanovitch

We comment on the preceding papers by Gerrard and Vernon concerning persuasion, perceived risk, and cancer-relevant behavior. Our purpose is to highlight several challenges for future investigators. First, relations between health cognition and health behavior (such as the link between perceived vulnerability and protective behaviors) are likely to be moderated by other variables, including individual differences and situational contexts. Second, we encourage health communication researchers to consider how persuasion is contextualized in social relationships and to employ mechanisms from the literature on social influence when promoting cancer prevention and early detection behaviors. Finally, we emphasize the importance of current feelings and anticipated emotions as motivators of salubrious actions. [Monogr Natl Cancer Inst 1999;25:119–22]

The effectiveness of interventions designed to promote cancer prevention and early detection behaviors often depends on the persuasiveness of a public service announcement, brochure, print advertisement, government letter, educational program, or communication from a health professional or cancer information specialist. From a psychological vantage point, these communications represent persuasion opportunities in which to apply a technology derived from nearly a half century of research on attitude change. The papers in the preceding section of this monograph suggest that applying this technology, however, might not always be a straightforward exercise. Sometimes the same communication strategies can produce opposite effects, depending on other contextual variables. Nowhere is this more clear than when trying to determine whether increasing a person's perceived vulnerability will motivate behavior change in a desirable way or will create unreasonable anxiety, defensiveness, and paralysis.

Affiliation of authors: Health, Emotion, and Behavior (HEB) Laboratory, Department of Psychology, Yale University, New Haven, CT.
Correspondence to: Peter Salovey, Ph.D., Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520-8205.

© Oxford University Press

Moderated Relationships

Our colleague, William J. McGuire, the dean of American attitude change researchers, is fond of saying, "The opposite of a great truth is also true" *(1)*. What does he mean by this mysterious and metaphysical Zen paradox? McGuire *(1)* urges us to recognize that important human behavior is not characterized best by "main effects." Rather, he asserts that the goal of science is the identification of specific moderators that help us to delineate the conditions under which a hypothesis is true and the conditions under which precisely the opposite hypothesis is true. It is unlikely, for example, that increasing perceived vulnerability always motivates desired health behaviors or that highly credible communicators are always more persuasive than those lacking in credibility. The search for these kinds of interactions—what neuroscientists like to call "double dissociations"—reveals the scope conditions under which various laws of human behavior hold or are violated.

The preceding papers were written in the spirit of McGuire's aphorism. Meg Gerrard and Sally Vernon admirably discuss studies in which the associations between psychological antecedents and desired health behaviors are moderated, making these relationships appear and disappear. Gerrard, for instance, reviews her work on self-esteem. Although it is usually assumed that individuals with high self-regard are less likely to engage in health-damaging behavior than those individuals who do not think highly of themselves, such is not always the case. In her work, young people with low self-esteem have a fairly realistic sense of how their sexual behavior places them at risk for sexually transmitted diseases (STDs), and they respond to new information about their vulnerability to STDs with appropriate concern; those individuals with high self-esteem respond to STD risk information more defensively. So self-esteem moderates the effect of risk communications in this domain. Gerrard also discusses the findings by Weinstein et al. *(2)*, showing that one's stage in the behavior change process moderates the effect of risk information on initiating salubrious actions. For example, information about the risk of lung cancer from household radon seems to motivate the purchase of test kits only among individuals who have not already decided on the value of buying them. Similarly, Vernon discusses the finding by Aiken et al. *(3)* that increasing perceptions of risk promotes screening mammography use among individuals with relatively few barriers preventing them from accessing this technology but not among those with many perceived barriers. These investigations, and numerous others, suggest that the relationship between vulnerability and cancer-relevant behavior is moderated by personal and situational variables that must be taken into account when considering the antecedents of relevant behavior change.

Persuasion: Past, Present, and Future

The importance of interactions in accounting for cancer prevention and early detection behaviors implies that the traditional approach to persuasion and attitude change, although valuable in delineating important "main effects," may be too restricted a perspective. Traditional psychological research on attitude change has focused primarily on one of three aspects of persuasive communication. These aspects have included (*a*) the source of the persuasive message (e.g., the communicator's expertise, credibility, trustworthiness, attractiveness, and similarity to the recipient), (*b*) the recipient of the message (e.g., the target's knowledge about the attitude domain, experience with the attitude object, and demographic and dispositional characteristics expected to be associated with influenceability), and (*c*) aspects of the message itself.

Of these three areas of research activity, message variables have been studied the least systematically, although interesting findings have emerged. For example, anecdotes about particular people are more persuasive than the presentation of cold statistics. Fear-arousing appeals are usually effective only when instructions about how to reduce the fear are included in the message. Forcefully delivered messages are more persuasive than quieter ones, and messages delivered quickly are surprisingly more effective than leisurely delivered messages [reviewed in *(4)*].

The two aspects of messages that have been studied most systematically in the context of cancer prevention and control are message framing and message tailoring. Message framing is a procedure derived from research on decision making motivated by prospect theory *(5–7)*. Message framing refers specifically to the emphasis in a message on the positive or negative consequences of adopting or failing to adopt a particular cancer-relevant behavior *(8)*. Appeals aimed at persuading individuals to perform cancer-relevant behaviors can be framed in different ways. In particular, information can be framed to emphasize relevant gains or losses. Gain-framed messages usually present the benefits that are accrued through adopting the behavior (e.g., "Obtaining a mammogram allows tumors to be detected early; this maximizes your treatment options"). Loss-framed messages generally convey the costs of not adopting the requested behavior (e.g., "If you do not obtain a mammogram, tumors cannot be detected early; this minimizes your treatment options"). Although these two messages convey essentially the same information, for certain behaviors one of these messages may be more persuasive than the other. In previous research *(9,10)*, we have found that messages concerning the value of using sunscreen to prevent skin cancer motivate appropriate sunscreen use when they are framed in terms of gains or benefits rather than losses or costs. However, we have also discovered that messages framed in terms of losses (costs) motivate more behavior change when the target behavior is screening mammography *(11)*. Thus, the message frame that is more effective depends on the type of health behavior being promoted—prevention or early detection.

Message tailoring is a procedure selected from the social-marketing arsenal. It refers to the adaptation of information and interventions that best fit the relevant needs and characteristics of specified target populations *(12)*. In this tradition, cancer-relevant messages have been tailored to the demographic characteristics of recipients, especially sex, ethnicity, occupation, and educational background, as well as to the stage in the behavior change process (reflecting prior contemplation of or experience with the targeted behavior). Some tailoring interventions create enormous libraries of text messages and accompanying graphics that are matched to complex and nearly unique combinations of previously identified recipient characteristics [e.g., *(13,14)*]. Recently, there has been increased interest in tailoring messages to recipients' psychological needs and personality characteristics [for a review, *see (15)*].

Although persuasion research using framing and tailoring strategies has certainly looked at scope conditions that moderate the relationship between the receipt of such messages and behavior change, future work in this arena will likely contextualize

framed and tailored messages in social relationships. That is, rather than focusing on the traditional source, message, and recipient of persuasive communications in isolation, the relationship between source and recipient will receive greater attention. This approach to persuasion is better characterized as social influence and is well captured in a framework developed by Robert Cialdini. Cialdini *(16)* describes six ways in which relationships between sources and targets play out in a persuasive context. These rules include

- *The power of authority:* We defer to authorities and even more so to mere symbols of authority.
- *Liking leads to influence:* We prefer to say yes to the requests of people we know and like.
- *Commitment, consistency, and cognitive dissonance:* Once we take a stand, we will encounter personal and interpersonal pressures to behave consistently with that commitment.
- *Reciprocation:* We try to repay, in kind, what another person has provided us.
- *Social proof:* We view a behavior as correct in a given situation to the degree that we see others performing it.
- *Scarcity:* Opportunities seem more valuable to us when they are perceived to be less available.

As health communication specialists, we need to consider the kinds of social environments that have facilitated persuasion and behavior change in other domains. Tupperware® storage containers have traditionally been sold by homeowners (rather than salespeople) to their neighbors in the context of household parties. Is it crazy to imagine "mammography parties"? After individuals donate blood, the Red Cross provides lapel buttons stating, "I gave blood today." Would it be too odd for clinics to provide buttons saying, "I received a mammogram today"? The men's clothing store in town has little trouble convincing customers to a buy a new tie or two after they have picked out an expensive new suit. Could a similar principle be employed to encourage sunsereen use among people who have made commitments to low-fat diets or smoking cessation? The idea here is to capitalize on social influence techniques that have worked in other domains and use them to promote health.

In considering persuasion campaigns for the purpose of cancer-risk reduction, capitalizing on these social psychological principles may increase the persuasive power of our brochures, educational programs, web sites, and the like. Organizing health communication around these kinds of ideas, however, strikes some people as, at best, manipulative and, at worst, unethical. The concern is that we are somehow tricking people into engaging in behaviors not based on their own rational calculation of the costs and benefits of adopting them. However, it is becoming increasingly clear that the rational consideration of impartially presented information and that the weighing of costs and benefits are not an especially accurate representation of what goes on in the minds of most decision makers when contemplating whether to use sunscreen, obtain a mammogram, or quit smoking.

In our own work, other mental processes seem to motivate health behavior, such as the emotional reactions that people anticipate feeling if they engage in the target behavior (or are blocked from engaging in that behavior). A friend recently installed a radon-mitigating fan in the basement of his house. This expenditure of about one thousand dollars came after testing his home's radon level several times and finding it a little above the recommended action level. His decision to pursue radon mitigation was not made because of a dramatically increased sense of lung cancer vulnerability. Rather, his concern was focused, primarily, on the regret and shame he would feel decades from now if he were to be diagnosed with lung cancer—regret and shame about his unwillingness to have spent a mere thousand dollars in his younger years. Mitigating household radon was a health behavior for this individual that served as an insurance policy against regret and shame in his twilight years. His actions were motivated only by a vague sense of perceived vulnerability but, more significantly, by anticipating future emotional reactions.

Conclusion

The communication of cancer-risk information and information about prevention and early detection behaviors represents an opportunity to change hearts, minds, and behaviors. In this paper, we have tried to emphasize several issues. First, relations between health cognition and health behavior (such as the link between perceived vulnerability and protective behaviors) are likely to be moderated by other variables, including individual differences and situational contexts. Second, we have indicated that the traditional focus in communication on source, message, and recipient variables has served us well but is rather limited. Instead, we encourage health communication researchers to consider how persuasion is contextualized in social relationships and to employ mechanisms from the literature on social influence when promoting health behavior. Finally, we emphasize the importance of affect—current feelings and anticipated emotions—as a motivator of health-relevant behavior. In a desire to preserve a view of the individual as a rational balancer of costs and benefits, we have created a false dichotomy between passion and reason [cf. *(17)*] and have underemphasized an important motivator for why individuals choose to protect their health.

References

(1) McGuire WJ. The yin and yang of progress in social psychology: seven koan. J Pers Soc Psychol 1973;26:446–56.

(2) Weinstein ND, Lyon JE, Sandman PM, Cuite CL. Experimental evidence for stages of health behavior change: the precaution adoption process model applied to home radon testing. Health Psychol 1998;17:445–53.

(3) Aiken LS, West SG, Woodward CK, Reno RR. Health beliefs and compliance with mammography-screening recommendations in asymptomatic women. Health Psychol 1994;13:122–9.

(4) McGuire WJ. Attitudes and attitude change. In: Lindzey G, Aronson E, editors. Handbook of social psychology. 3rd ed. New York (NY): Random House; 1985. p. 233–346.

(5) Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979;47:263–91.

(6) Kahneman D, Tversky A. The psychology of preferences. Sci Am 1982; 247:160–73.

(7) Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981;211:453–8.

(8) Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997;121:3–19.

(9) Detweiler JB, Bedell BT, Salovey P, Pronin E, Rothman AJ. Message framing and sunscreen use: gain-framed messages motivate beach-goers. Health Psychol 1999;18:189–96.

(10) Rothman AJ, Salovey P, Antone C, Keough K, Martin CD. The influence

of message framing on intentions to perform health behaviors. J Exp Soc Psychol 1993;29:408–33.

(11) Banks SM, Salovey P, Greener S, Rothman AJ, Moyer A, Beauvais J, et al. The effects of message framing on mammography utilization. Health Psychol 1995;14:178–84.

(12) Pasick RJ. Socioeconomic and cultural factors in the development and use of theory. In: Glank K, Lewis FM, Rimer BK, editors. Health behavior and health education: theory, research, and practice. San Francisco (CA): Jossey-Bass; 1997. p. 425–40.

(13) Marcus AC, Heimendinger J, Wolfe P, Rimer BK, Morra M, Cox D, et al. Increasing fruit and vegetable consumption among callers to the CIS: results from a randomized trial. Prev Med 1998;27:S16–28.

(14) Skinner CS, Siegfried JC, Kegler MC, Strecher VJ. The potential of computers in patient education. Patient Educ Couns 1993;22:27–34.

(15) Rimer BK, Glassman B. Tailoring communications for primary care settings. Methods Inf Med 1998;37:171–7.

(16) Cialdini RB. Influence: science and practice. 3rd ed. New York (NY): HarperCollins; 1993.

(17) Salovey P, Mayer JD. Emotional intelligence. Imagination Cogn Pers 1990; 9:185–211.

Introduction of Section: Implications for Improving Risk Communication Through Various Channels

J. Paul Van Nevel

The papers in this section of the monograph provide an overview of the lessons learned and research opportunities ahead for cancer risk communication in very practical health care settings and through different types of media and presentation formats. These papers examine the role of cancer risk communication in the clinic; in some of the newer, more interactive media; and in the more traditional broadcast and print media. Each of the authors is doing work that is setting new standards for risk communication.

Communicating messages that include risk information is a difficult but essential task. As director of the Office of Cancer Communications of the National Cancer Institute (NCI), I work with a staff of media professionals, science writers, health educators, and health communicators to translate the science about cancer risk and to communicate it in an understandable and useful manner. We often are confounded by the many challenges and barriers to doing this effectively: Audiences vary greatly; the information often has a degree of uncertainty; sometimes a risk factor has both advantages and disadvantages, as with tamoxifen (*see below*); and preconceived notions among target audiences influence how the message is received and how the audiences act on the information. In addition, communication of a research finding or of a health message dealing with a risk factor is often complicated by communications about the same risk by competing interests. Environmental advocates and industry often take opposite stands on a risk factor, and the public is bombarded with conflicting messages.

On a practical basis, the need to communicate cancer risk accurately is ever present, and there are at least three major ways in which this need is presented to the communicator.

First, the need for sueh communication is well known sometimes and does not present itself unexpectedly. An example is presentation of risks of tobacco use to teenagers. But we need more effective ways to do this.

Second, sometimes a risk situation, while sudden and unexpected, presents a need to communicate quickly through mass media and then to follow up with risk communication that is more planned. For example, when a large clinical trial showed that tamoxifen was effective in reducing breast cancer risk, it was clear that the NCI needed to use the mass media to inform the public of the benefit of this compound in reducing risk and also of its potential risks, particularly to subsets of women who would use it. In addition, the situation demanded a long-range risk communication program aimed at health professionals and women who need to make individual decisions about whether to use tamoxifen to reduce breast cancer risk. As a result, the NCI developed and distributed a Breast Cancer Risk Assessment Tool. This computerized program, later adapted for use on the Internet, is used by health professionals with their patients to estimate a woman's risk of developing breast cancer over the next 5 years and over her lifetime and thereby helps inform a decision about whether to take tamoxifen. In light of tamoxifen's potential side effects, the NCI—at the time of this conference—was considering adding to the tool information on the risk of side effects, even though estimates of an individual's risk of these side effects would be less certain than estimates of her risk of breast cancer.

Third, sometimes the need to communicate risk is sudden and unexpected but does not last long. These episodes often occur after a report in a journal that a particular compound causes cancer or protects against it. Such episodes are over quickly, with mass media the main channel for communication, because these reports do not have enough of a scientific basis, or do not cause enough of an unfounded public furor, to ever become the subject of planned risk communication.

When presented with a need to communicate cancer risk clearly, how should it be done? It is clear that no single method will work in isolation. The papers that follow, by Schwartz and Woloshin, Strecher et al., Rimer and Glassman, and Lipkus and Hollands, outline a number of innovative and complementary methods of communicating cancer risk information. The final three papers, by Levine, Brody, and Russell—all journalists—examine the use of mass media, both broadcast and print, to communicate risk information. These three papers reflect lessons learned through many years of experience more than they reflect the outcomes of specific studies.

The mass media represent an incredibly powerful tool for communicating health information. Survey after survey has shown that people receive a great deal of health and risk information from the mass media. Unfortunately, information presented in these outlets is uncontrolled and very unlike more controllable situations, such as planned ventures, including one-on-one or one-on-small-group interactions. The mass media tend to deal with episodes of science that are positioned as news. These episodes spring from reports in medical and scientific journals or from community events, such as the occurrence of cancer clusters that get suddenly noticed and reported in the press. Because these news reports are prepared quickly, they frequently are reported out of context and can create a false impression. If a journal publishes a study that associates coffee consumption with pancreatic cancer, for example, it probably will be reported in the mass media that coffee causes cancer, and the report will not include the context of a larger body of literature that substantiates or refutes the findings of a single study.

Mass media reports are often shaped by advocates who seek either to minimize or to maximize the risk associated with something. Alar and saccharin are examples. Often it is the industry on one side, environmental activists on the other, and a confused public in the middle.

Affiliation of author: Office of Cancer Communications, National Cancer Institute, Bethesda, MD.

Correspondence to: J. P. Van Nevel, National Institutes of Health, Bldg. 31, Rm. 10A31, MSC 2580, Bethesda, MD 20892 (e-mail: nevelp@occ.nci.nih.gov).

Also, mass media use human anecdotes to bring a story to life for a reader, viewer, or listener, and the drama and emotion of the anecdotes can distort and mislead. Mass media reports can also be shaped by the scientists who did the research that resulted in news. How do these scientists couch the results? Do they give them more certainty or more importance than they should have? There is a temptation for both individuals and for institutions to exaggerate results one way or another to reflect individual or institutional interests, thereby resulting in distortion or inaccuracy of message.

The papers in this section explore ways to improve risk communication through the mass media and other channels, and they include suggestions that can be put into practice immediately or which show us the way of the future. The section ends with a discussion of the seven papers by Fong and Russell.

Risk Communication in Clinical Practice: Putting Cancer in Context

Lisa M. Schwartz, Steven Woloshin, H. Gilbert Welch

Context: **Clinicians are increasingly urged—even mandated—to help patients make informed medical decisions by paying more attention to risk counseling. For many, the role of risk counseling is new and unfamiliar. This effort is made more difficult given the practical constraints created by 15-minute visits and competing demands (e.g., patient's chief complaint and institutional needs).** ***Objective:*** **We detail a three-part approach for improving risk communication, acknowledging the role of clinicians, patients, and other communicators (i.e., media or public health agencies).** ***Proposed Approach: Office-based tools to help clinicians do more.*** **We suggest two ways to help make up-to-date estimates of disease risk and treatment benefit easily available during office visits. We propose the development of a comprehensive population database about disease risk and treatment benefit to be created and maintained by the federal government.** ***Educating patients.*** **We propose "Understanding Numbers in Health" a tutorial that reviews basic concepts of probability and their application to medical studies to help people become better critical readers of health information.** ***Guidance for communicators.*** **Finally, we propose a writer's guide to risk communication: a set of principles to help health communicators present data to the public clearly and objectively.** ***Conclusion:*** **In addition to tools to help clinicians better communicate risk information, serious efforts to improve risk communication must go beyond the clinic. Efforts that help the public to better interpret health risk information and guide communicators to better present such information are a place to start. [Monogr Natl Cancer Inst 1999; 25:124–33]**

Introduction

Clinicians are increasingly urged [even mandated *(1)*] to do better risk counseling to help patients make informed medical decisions. The central counseling role that clinicians are expected to play is typified by the recent National Institutes of Health (NIH) Consensus Panel on breast cancer screening for women ages 40–49 that stated that "... a woman should have access to the best possible information in an understandable and usable form. Her health care provider must be equipped with sufficient information to facilitate her decision-making process" *(2)*. Unfortunately, the high expectations articulated in this and similar recommendations rarely are accompanied by practical advice.

There are several reasons why efforts focused on the clinician may have limited effect in improving risk communication. First, for many clinicians, risk communieation is an unfamiliar discipline. The emphasis on the patient's role in medical decisions is a relatively recent phenomenon (i.e., shared decision making) *(3)*. Few clinicians reeeive training in methods to promote effective eommunication with patients (about risk or any subject), and, in fact, little is known about the best ways to communicate such information. Moreover, the relevant data to be communicated have only recently become available and are not easily accessible at the time of office visits.

Second, the competing demands of clinical practice limit what clinicians can be expected to do within the real-world constraint of the standard 15-minute visit *(4)*. The clinician first must address the patient's chief complaint, the concerns and symptoms that brought the patient to the office (which is usually not "I need help to make sense of the health risks I face"). In addition, the growing institutional demands aimed at measuring and at improving the quality of health care have already left many clinicians feeling burdened. These demands invariably result in increased paperwork for data collection and for monitoring adherence to practice guidelines. Given the foregoing, it is not surprising that the limited data available suggest risk is rarely discussed in typical clinical encounters *(5)*.

Risk communication is particularly important in discussions about cancer. Because cancer is an especially dreaded diagnosis, information about the chance of developing cancer or the effect of various preventive strategies in reducing cancer risk or the chances of dying of cancer may be extremely welcome. However, information about cancer alone (or a particular cancer) may overemphasize the risk compared with other health issues. In this paper, we focus on ways to improve the presentation and

Affiliations of authors: Department of Veterans Affairs Medical Center, White River Junction, VT; Center for Evaluative Clinical Sciences, Dartmouth Medical School, Hanover, NH; Norris Cotton Cancer Center, Lebanon, NH.

Correspondence to: Lisa M. Schwartz, M.D., M.S., and Steven Woloshin, M.D., M.S., VA Outcomes Group (111B), VA Medical & Regional Office Center, 215 N. Main St., White River Junction, VT 05009.

See "Notes" following "References."

interpretation of quantitative data about risk in general (Table 1). First, we make suggestions for simple office-based tools to help clinicians communicate about prevention. Next, we discuss a strategy for educating patients to be better consumers of data. We conclude with guidance for communicators to improve the quality of data disseminated to the public by news media and public health agencies.

Office-Based Tools: Helping Clinicians Communicate the Value of Prevention

Rationale

A fundamental goal of health risk communication is to help people better understand the important health risks they face. This goal, a basic concept of contemporary medical ethics *(3)*, also has practical implications. Patients who received more information from their physician were more satisfied and had higher compliance with medical regimens *(6)*. At a minimum, understanding the magnitude of a risk (i.e., how big of a threat is breast cancer to me?) entails having some idea of what the risk is (what does it mean to have breast cancer?) and the chances of developing or dying of the condition. Although it is often assumed that physicians spend much time communicating with their patients about risk, remarkably little is known about if and how such communication actually occurs. In the one study *(5)* documenting doctor–patient risk communication (defined as discussion about behavior change, compliance with screening tests, or preventive treatments), risk was discussed in only 26% of primary care visits and was described numerically in only 3%. One reason why physicians may not engage in risk communication with patients is that they lack easy access to the relevant data. Simple office-based tools may help overcome this barrier.

Office-based tools may be of value in stimulating and in facilitating discussions about disease risks. Patients may want to know the answer to questions such as, what is the chance that a person my age will die of heart disease or breast cancer in the next 10 years? Similarly, patients may also find information about the benefit of various risk-reducing strategies valuable: for example, how does my chance of dying of breast cancer change if I have annual mammograms? To be useful, such office-based tools need to be up to date, immediately available, and easy for both clinicians and patients to use and to understand.

Disease-Specific Tools

A number of tools that generate disease-specific risk estimates for an individual patient are now available. For example, the American Heart Association *(7)* has a web site where an individual's risk of myocardial infarction can be calculated with the use of a model generated from the Framingham data. The Northern New England Cardiovascular Group *(8)* uses a preprogrammed hand-held computer to provide patients considering coronary artery bypass graft surgery with an estimate of the mortality risk they face from surgery. Recently, the National Cancer Institute (NCI) *(9)* issued the Breast Cancer Risk Assessment Tool that provides women with their risk of developing breast cancer to help women contemplating tamoxifen for the primary prevention of breast cancer.

Implementing these tools in clinical practice entails collecting the necessary risk factor information from patients [e.g., the breast cancer risk factors required for Gail model *(10)* could be collected before a clinic visit] and generating a risk report. Such risk reports could then be attached to each patient's chart at the time of a scheduled clinic appointment with their clinician to maximize the chance of discussion. Some evidence *(11–13)* suggests that such personalized messages may be more effective than generic messages. Whether the extra time, cost, and technical difficulty of these personalized reports outweigh this potential advantage is unknown.

Although such tools are appealing because the disease-specific estimate is personalized, the inherent focus on a single disease taken out of context may overweigh its importance. When making a decision, a patient may find it helpful to understand where this particular disease fits into the important health threats he or she faces. Patients may find it particularly helpful to know: How does my chance of dying of this particular disease compare with other diseases? What is my overall chance of dying? How does the overall mortality benefit of one intervention (e.g., mammography) compare with the benefit of another (e.g., giving up cigarettes)?

Table 1. Proposed three-part approach to improve health risk communication*

Component	Problem	Proposed approach
Clinical encounter	Physicians infrequently discuss risk with patients and rarely use quantitative terms in such discussions	**Office-based tools** Help clinicians to do more • Disease-specific tools Collection of risk factor data to generate personalized disease risk estimates for clinic visit (e.g., NCI Breast Cancer Risk Assessment Tool) • Comprehensive tools Wall charts with age- and sex-specific population data about disease risk and benefit of interventions (Tables 2–3)
Patient comprehension	Low levels of numeracy and susceptibility to framing are important barriers to patient comprehension	**Patient education** Teach patients to be better consumers of data • General education Primer to teach patients about numbers in health and how to rate the quality of scientific evidence
Media messages	Misleading messages contribute to unrealistic beliefs about treatment benefit	**Guidance for communicators** Improve quality of data patients see • Set of guiding principles (Table 4)

*Each approach will require evaluation for feasibility and effectiveness. NCI = National Cancer Institute.

Comprehensive Tools: Disease Risk and Benefit Wall Charts

To provide this context, we propose the creation of charts with age- and sex-specific data about disease risks and treatment benefits. Tables 2 and 3 present examples of such simple office tools. Such low-tech tools, although lacking the glamour of interactive computer applications, have several distinct advantages. Simple tools are inexpensive and could be used anywhere (e.g., posted in any clinic office). Furthermore, simple tools require no special hardware and no additional personnel or maintenance.

Disease Risk Chart

The disease risk chart shown in Table 2 displays 10-year disease-specific mortality data for five major diseases—in this case, coronary artery disease, breast cancer, lung cancer, colorectal cancer, and ovarian cancer—for women within 5-year age categories *(14,15)*. Moving across the table allows the user to compare the magnitude of each disease risk. Because many people may be even more concerned about their overall chance of dying, the final column displays all-cause mortality to provide context about how much each disease contributes to the overall chance of dying. Whereas mortality data can be represented as counts, proportions, or rates, we use counts with a stable denominator (e.g., 100 000 women) in our example, because there is some evidence *(16,17)* suggesting that people find counts easiest to understand.

Benefit Chart

Table 3 presents an example of a benefit chart. The goal of this chart is to help patients compare the relative effect of a change in behavior or specific intervention on all-cause mortality. Our example displays age- and sex-specific 10-year all-cause mortality with or without a given intervention. The numbers shown in the chart are crude estimates that are accurate in terms of order of magnitude. The first scenario in the chart considers 100 000 smokers and displays their chance of dying in the next 10 years if they all continue to smoke or if they all quit smoking and the net effect—about 6500 deaths prevented among 55–59 year-old smokers. Another scenario considers 100 000 women who do not have an annual screening mammography and those who do and shows the net effect of 200 deaths prevented for 55–59 year-old women. These examples show that, for a 55-year-old female smoker, giving up cigarettes has a substantially greater effect on all-cause mortality than annual mammography.

Ideally, we would create a benefit chart for an intervention only if the efficacy of screening or behavioral changes has been demonstrated in randomized trials (e.g., mammography for women in their 50s) or when observational analytic studies have convincingly demonstrated benefit and the interventions are routinely recommended (e.g., Pap screening for cervical cancer, smoking cessation). Because age and comorbidity (i.e., competing risks of disability or death that patients face in addition to the risk under consideration) will importantly influence the benefit of any intervention (behavioral changes or screening tests), the benefit charts may encourage explicit discussion between the patient and clinician about these issues. Because interventions can also have harms, an important challenge remains in how to convey data about side effects, bad outcomes, and so forth. Studies comparing the effectiveness of our proposed comprehensive tools, disease-specific tools, and usual care are needed to learn which better helps patients make important medical decisions.

Data Source for Charts

The data required to construct such charts are currently available from a variety of sources [e.g., statistical abstracts, the

Table 2. Disease risk chart for women*: estimated 10-year disease-specific and all-cause mortality

	Imagine 100 000 women your age. Over the next 10 years, how many will die of					
	Coronary disease†	Lung cancer	Breast cancer	Colorectal cancer	Ovarian cancer	All causes
For women age (y)						
20–24	8	1	6	2	2	600
25–29	30	5	30	6	5	700
30–34	70	20	70	10	10	1000
35–39	140	50	150	30	20	1500
40–44	300	130	270	50	50	2100
45–49	630	310	420	100	90	3300
50–54	1200	600	550	180	150	5100
55–59	2200	1000	680	300	210	8100
60–64	3900	1500	830	440	280	12 000
65–69	6500	1800	970	640	350	18 000
70–74	11 000	2000	1100	880	400	27 000
75–79	18 000	1900	1200	1200	440	41 000
80–84	34 000	1500	1200	1500	400	67 000
85+	42 000	940	1100	1500	300	79 000

*We obtained 1996 mortality rates for 5-year age groups from Surveillance, Epidemiology, and End Results Program (cancer rates) and National Center for Health Statistics (coronary disease, all-cause mortality). We converted these annual rates into 10-year probabilities by applying the age-specific, disease-specific, and all-cause mortality rate for adjacent 5-year intervals. Numbers have been rounded to facilitate interpretation and represent rough estimates that should be accurate in terms of order of magnitude.

For example, about 8 of 100 000 women age 20–24 will die from coronary artery disease in the next 10 years compared with 3900 out of 100 000 women age 60–64.

†Coronary artery disease includes deaths attributed to acute myocardial infarction, old myocardial infarction, angina, ischemic heart disease, and subacute/acute forms of ischemic heart disease.

Table 3. Prototype benefit chart for women: 10-year all-cause mortality with and without intervention and deaths prevented (95% confidence interval)

Imagine 100 000 women your age. Over the next 10 years, how many will die if they									
	Quit smoking			Start annual screening with mammography			Start annual screening for colon cancer*		
Age (y)	No	Yes	Deaths prevented	No	Yes	Deaths prevented*	No	Yes	Deaths prevented*
55–59	13 000	6500	6500 (5000–8000)	8100	7900	200 (100–300)	8100	8000	100 (90–110)

*These numbers assume that the reduction in disease-specific mortality observed in trials is extended to all-cause mortality. Reductions in all-cause mortality have not been consistently observed in these screening trials. These estimates, therefore, represent best-case scenarios.

National Center for Health Statistics, and the Surveillance, Epidemiology, and End Results (SEER) Program[1]] but would be difficult to consolidate and update. A health risk database could be developed, maintained, and made publicly available by the federal government. Such a central repository of risk information would serve the public good in much the same way as Statistical Abstracts of the U.S. or other federally maintained databases. A distinct advantage of a federal agency taking on this responsibility would be to minimize the incentives to advocate for a specific disease. The National Center for Health Statistics already collects disease data and would be an ideal candidate for the disease risk chart. Because benefit data would require more critical interpretation of the literature, the Agency for Health Care Policy and Research, with its interest and expertise in evidence-based medicine, would be a natural choice for this responsibility.

Patient Education: Teaching Patients to Be Better Consumers of Data

Rationale

Efforts to promote informed patient decision making have become increasingly common. In general, these efforts have focused on providing disease-specific facts. The rationale underlying this approach is straightforward: to make informed decisions requires information. If people lack key facts, their decisions cannot be informed. The solution, then, is to provide the facts.

Unfortunately, there are reasons to question the likely effect of this commonsense approach to patient education. First, patients may not be ready for the data. That is, problems with numeracy (i.e., low quantitative literacy) are common. For example, in the National Adult Literacy Survey *(18)*, 47% of adult Americans could not calculate the difference between regular and a sale price from an advertisement. Low levels of numeracy strongly relate to difficulty in making use of quantitative data about the risk reduction of screening mammography *(19)*. Second, patients may not know how to interpret the information they are given. Educators have long understood that presenting facts without first preparing the audience to receive them (i.e., integrate them into some organizing structure) is ineffective and probably counterproductive. In such a case, the members of the audience will absorb little information (which will be quickly forgotten), will not understand how the information fits into their own experience, and may misinterpret what it means. With little experience in using data, for example, patients may be especially susceptible to the framing effects frequently discussed in the cognitive psychology literature (i.e., how simple changes in the format of otherwise identical numerical information can profoundly influence its interpretation) *(20–26)*.

To see how well patient educational materials convey quantitative data to patients under the best of circumstances, we performed a structured literature review *(27)* to identify randomized trials of interventions designed to communicate quantitative data about disease risk or treatment benefit. Of 70 trials studying patient education interventions, only four attempted to provide patients with some sort of framework for approaching a medical encounter (although none dealt with the interpretation of quantitative data). The rest presented facts without any interpretative framework. Whereas the majority of these trials sought to convey quantitative data (n = 47 articles), we found only seven randomized trials evaluating patient comprehension of these data *(19,28–33)*. The table in the Appendix summarizes the results of five of the seven trials that tested a patient education tool *(28–32)*. Although it is difficult to compare across studies because the interventions and metrics of efficacy are quite diverse, the interventions had variable effects and, in general, left substantial room for improvement.

Tutorial: Understanding Numbers in Health

Rather than relying on clinicians and communicators to interpret information for the public, we propose to develop the public's capacity to be critical consumers of health information—to prepare patients to receive data. Our proposal consists of creating a generic patient's user guide to health information that deals with the following five subjects.

1) *What is risk?* Attempts to discuss medical risk are easily undermined by confusing and imprecise use of language. The tutorial begins by addressing common sources of confusion (multiple meanings of the word "risk"), how to use words (and the limits of words) in describing risk, and ways to quantify risk (probability, percents, proportions, and rates). We will also introduce the reader to a scale that we have developed to facilitate quantification and communication of risk—particularly small risks less than 1% (Fig. 1).

2) *What to look for in a statement about risk.* This section teaches the reader to look for various essential elements in statements about risk. Readers will be sensitized to ask questions, such as: What is the risk under discussion (e.g., is it the risk of being diagnosed or of dying?)? What is the time frame under consideration (e.g., next 5 years or lifetime—and what does "lifetime" mean?)? Who is at risk (i.e., does the

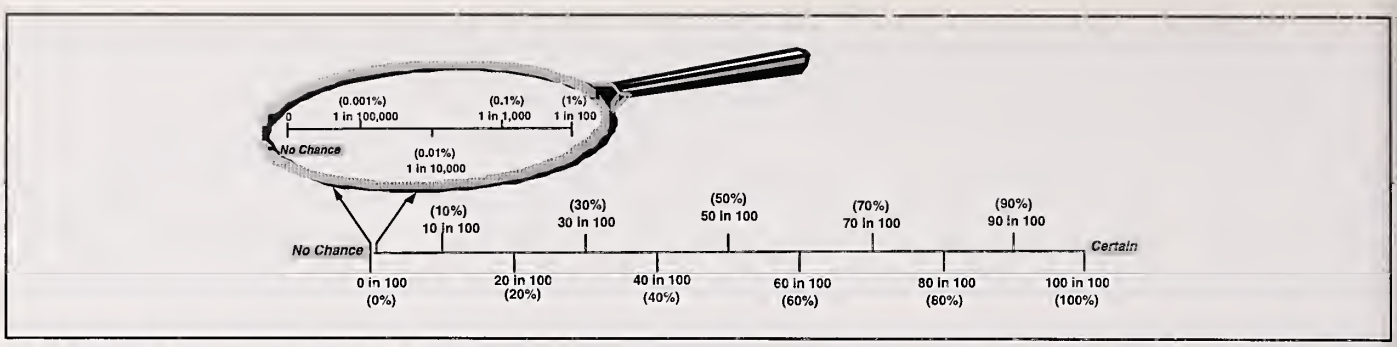


Fig. 1. Visual scale for representing event probability.

statement refer to all women? women of a certain age? women with specific characteristics such as a family history of breast cancer?)?

3) *Putting risk in context.* A salient but rare outcome, such as a celebrity dying of a rare cancer, may give undue weight to certain health risks. The tutorial emphasizes the need to put risks into perspective. Readers will be encouraged to ask questions, such as: How does the chance of this disease compare with other diseases or other familiar events? How dangerous is the disease (i.e., appreciating the difference between developing a condition and dying of it)? To illustrate competing risks, we will make use of disease risk charts discussed previously.

4) *Changing risk.* This section focuses on how to interpret statements that measure changes in risk given some exposure or intervention (e.g., relative and absolute risk reduction or number needed to treat) and introduce the concept of framing (e.g., dying versus not dying). Benefit charts could be used to highlight that not all risk factors and interventions are equally important.

5) *Evidence.* The final section points out that there is uncertainty in what we know and introduces the idea of grading evidence by highlighting basic concepts of study design (e.g., observational study versus randomized clinical trial). We encourage readers to have a healthy skepticism and ask themselves, "Can I believe what I am being told? Could it be wrong?"

Limitations

Our approach has several potential problems that should be acknowledged. First, some patients say they do not want information. Many of these patients would therefore have no interest in our tutorial. It is possible that for some people, however, an expressed lack of interest in information may really mean they are afraid they will not be able to understand what they are told. Our tutorial may make quantitative information accessible to people who might otherwise have given up. Next, patients' interest in the tutorial may change substantially under different circumstances. For example, it is possible that patients facing new and serious decisions (e.g., at the time of learning a new cancer diagnosis) may feel too emotionally overwhelmed to make use of the materials, whereas patients making decisions under less intense circumstances (e.g., an older man considering prostate-specific antigen screening) may find the materials especially useful. Finally, understanding whether the tutorial succeeds in teaching (i.e., what do patients learn?) and whether the materials help people make decisions will require careful study and will be the subject of future research.

Application

Assuming that we are able to demonstrate that the tutorial is usable, acceptable, and effective in a randomized trial, we could envision the tutorial being used in a variety of settings. The tutorial could be available for use in school curricula (i.e., modeled on "Chance," an Internet-based, quantitative literacy course that is based on current chance events in the news and is available at: http://www.dartmouth.edu/%7Echance/Chance.html). In clinical settings, the tutorial could be distributed as part of general patient orientation to a practice or could be distributed at the time that important decisions are being made (e.g., men newly diagnosed with prostate cancer or women newly diagnosed with breast cancer deciding on a treatment course) and interpretation of data becomes critical.

Guidance for Communicators: Improving the Quality of Data That Patients See

Rationale

Communicators face the difficult task of translating—often under short deadlines—complex, probabilistic information into a format accessible to a general public with limited grounding in science and with limited ability to make use of probabilistic information *(18,19)*. Communicators themselves may have only superficial training in the critical evaluation of medical literature. They may be unprepared to recognize potential biases, methodologic weaknesses, or questionable statistical manipulations that ought to raise caution about the validity or generalizability of a study's results. For example, the case for cancer screening is often made with a statement that the 5-year survival of patients diagnosed with early stage cancers is much greater than that of patients diagnosed with later stage cancers. These inherently biased statements do not demonstrate that screening is beneficial. Rather, these statement simply say that patients diagnosed earlier live with a cancer diagnosis longer *(34,35)*. It is only from the results of randomized trials that demonstrate that those who are screened have lower cancer death rates than those who were not screened that we can know the true effect of screening.

Evaluation of the accuracy of communications targeting the general public are limited, but frequent problems with news reports have been demonstrated *(36,37)*, and a recent review *(38)* of Australian public health brochures about screening mammography documented unbalanced and incomplete presentation of data, suggesting an underlying attempt to persuade rather than inform.

Guidance for Communicators

Table 4 presents a set of principles that we hope will guide communicators in how to present the data completely, objectively, and understandably. To illustrate some of these principles, we will use examples from the NCI's "Breast Cancer Risk Tool: An Interactive Patient Education Tool" *(9)*.

Delineate the Main Message Clearly

Breast cancer risk. Fig. 2 shows the NCI screen that displays the main message about breast cancer risk. A woman's Gail risk factors are entered, and then her calculated breast cancer risk is displayed graphically and described in text as follows:

> "Estimated risk for invasive cancer over the next 5 years is 0.6%"

> "Estimated risk for invasive cancer over her lifetime is 11.1%"

The NCI tool has done well in clearly defining this main message: the outcome under consideration is clearly stated as the 5-year and lifetime (to age 90) risk of a diagnosis of invasive breast cancer.

Table 4. Guiding principles for communicators

Delineate the main message clearly
Define the outcome under consideration
Diagnosis, specific morbidity, or death from disease
Provide the time frame (e.g., in the next 10 years)
Present data clearly
Some formats are hard to understand and should be avoided
e.g., "1 in ____" or expressing small risks with percentages alone (i.e., 0.01%)
Better formats use counts and balanced framing
e.g., Imagine 1000 people: 100 people will die from cancer, 900 people will not
Clearly specify to whom the data apply (e.g., gender, age, risk factors)
Present benefit and harm symmetrically
When expressing changes in risk, present absolute event rates or absolute changes from baseline risk
Provide context
Present both chance of diagnosis and death to reflect disease lethality
Specify important competing risks for death
Disease under consideration may be a less important contributor to a person's overall chance of dying than other diseases (e.g., prostate cancer versus heart disease for 75-year-old man)
Compare with familiar events
Help to calibrate reader to risk magnitudes by providing chance of events generally acknowledged to be common (e.g., catching a cold) or extremely rare (e.g., being hit by lightening)
Benchmark factor (risk factor or intervention) under consideration against other known factors to be clear that all factors do not change risk by the same amount
Acknowledge uncertainty
Be explicit about uncertainty by presenting some measure of estimate precision (e.g., 95% confidence interval)
Extrapolate cautiously from a single study, intermediate end point (e.g., tumor shrinkage), or across populations; express grater caution about inferences from weaker data (e.g., observational study)
In the special case of screening, remember that the benefit can only be shown in randomized trials; improved 5-year survival for cancers detected by screening is not proof of benefit
Remember health
Scary messages do not make people feel healthier and may generate unrealistic expectations about disease risk and treatment benefit

Risk is expressed as a percentage in text and on a linear percentage scale (i.e., 0%–50%, marked with 5% increments). This dual presentation is a particularly good idea because many people have trouble working percentages alone *(18,19)*, especially percentages less than 1%. For example, only 20% of female veterans—almost all of whom had graduated from high school—were able to correctly convert 0.1% to 1 in 1000 *(19)*. There is surprisingly little guidance available on how best to present such quantitative information. Some prior studies *(16,17)* suggest that counts (e.g., imagine 1000 women, 10 die) may be easier to understand than percentages, and, in a recent study *(39)*, we demonstrated that people have great difficulty with expressions of the form "1 in ____." Unfortunately, the design of the graphic in Fig. 1 is not ideal. It is practically impossible to indicate probabilities below 1% (a relevant range for many likely users). One approach to this problem is to use a scale, like the one shown in Fig. 1, designed to facilitate expression of small probabilities.

Benefit of tamoxifen. The NCI tool has a second main message that is to inform women about the benefit of tamoxifen in the primary prevention of breast cancer. This message is less well done.

> "Women [taking tamoxifen] had about 49% fewer diagnoses of invasive breast cancer"

The benefit of tamoxifen is only expressed as a relative risk reduction without an explicit statement about baseline risk. Several studies *(20–23)* have shown that physicians and patients find the benefit of an intervention more compelling when it is expressed as a relative risk reduction rather than the corresponding absolute risk reduction. Whereas most typical risk reduction expressions may be difficult to understand, the relative risk reduction without the baseline risk—the format used in the NCI tool—is particularly difficult *(19)*. In the Breast Cancer Prevention Trial *(40)*, the baseline risk (the chance per year of developing invasive breast cancer for women in the placebo arm) was 68 cases per 10 000 women per year. Applying the 49% relative risk reduction yields a risk of 34 cases per 10 000 women per year in the tamoxifen arm.

Curiously, one of the more salient potential harms of tamoxifen—an increased chance of developing uterine cancer—is presented using absolute event rates for each group:

> ". . . annual rate of uterine cancer in the tamoxifen arm was 30 per 10 000 compared to 8 per 10 000 in the placebo arm"

This asymmetric framing tends to emphasize the benefit of tamoxifen while minimizing the harm (Fig. 3). If the increased uterine cancer was expressed using the relative risk format, the statement would read ". . . 275% more uterine cancer" and would likely elicit a very different feeling. On the basis of this framing phenomenon, we believe it is important to present both the benefits and harms of a treatment using the same frame. To enhance the effectiveness of such messages, we suggest that communicators present changes in risk using absolute event rates *(19)*.

Provide Context

The purpose of the NCI tool is "to measure a woman's risk of invasive breast cancer." The risk provided is the risk of *getting*

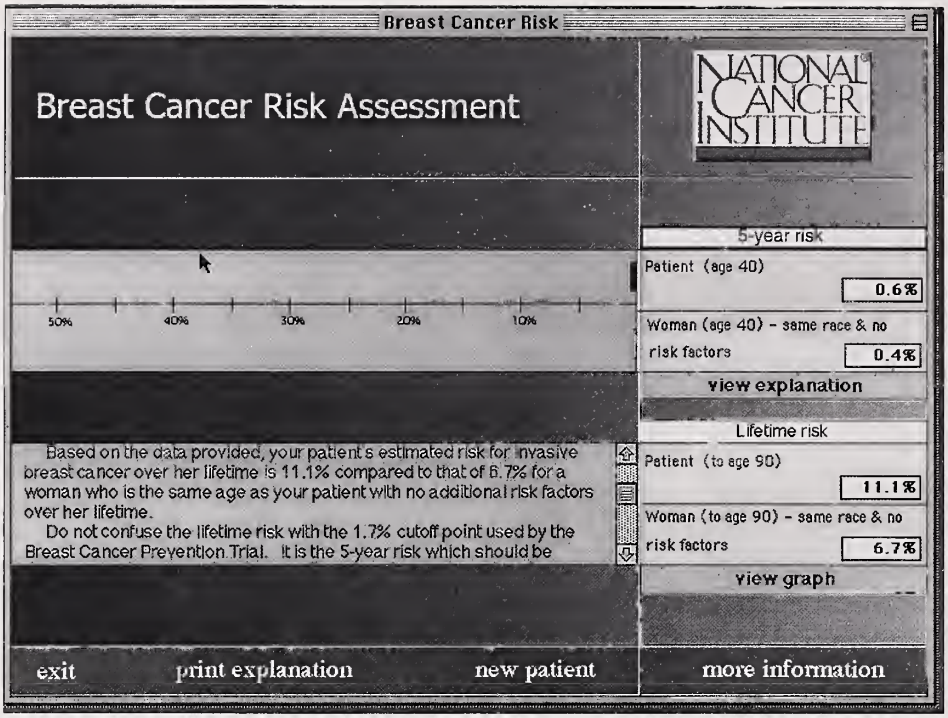


Fig. 2. The National Cancer Institute's Breast Cancer Risk Assessment screen displaying a woman's 5-year and lifetime risk of invasive breast cancer.

	Benefit Breast cancer diagnosis	Harm Uterine cancer diagnosis
Actual "frames" used in NCI risk tool	49% reduction	Placebo: 8 per 10,000 versus Tamoxifen: 30 per 10,000
Alternate "frames"	Placebo: 68 per 10,000 versus Tamoxifen: 34 per 10,000	275% increase

Fig. 3. Asymmetric presentation of benefit and harms of tamoxifen. The actual frames tend to emphasize benefit and minimize harm, whereas the alternate frames have the opposite effect. The numbers shown are the average annual event rates.

breast cancer. For many women, however, the more relevant risk is her chance of *dying* of breast cancer. Presenting incidence data without mortality data fails to provide important context about how often breast cancer results in death. A related issue involves competing risks for death—for example, how a woman's chance of dying of breast cancer compares with her chance of dying of heart disease.

Another important aspect of context relates to calibrating users to the magnitude of the probabilities presented. It has been demonstrated that even experts are often poorly calibrated to the magnitude of various risks *(41)*. Providing comparisons with the chance of familiar events—such as having a minor car accident—may help make the numbers more meaningful. Such comparisons might help users put their breast cancer risk (i.e., numbers like 0.6% given above) into perspective.

In addition, when discussing factors that change risk, it is important to emphasize that all risk factors do not increase risk to the same degree—something that is not done in the tool. For example, a woman may be able to better judge her breast cancer risk by knowing that family history and age raise the chance of breast cancer to a far greater extent that having the onset of menarche at an early age.

Acknowledge Uncertainty

In both the presentation of disease risk and treatment benefit described above, only point estimates (e.g., 49% risk reduction) are provided. These single numbers without 95% confidence intervals imply a false sense of certainty in the expected outcome. This concern is mostly relevant to discussion of treatment benefit rather than disease risk. Whereas the formal statistical definition of 95% confidence intervals may be diffi-

cult to communicate, we suggest a simple statement that uses the lower and upper bound of the confidence interval in the following way:

> "If 1000 woman do not take tamoxifen, six will be diagnosed with invasive breast cancer in the next year. If these 1000 women all take tamoxifen, our best guess is that three of these six women will **not** get breast caneer. It is possible that tamoxifen actually prevents as few as two women or as many as four women from getting breast cancer."

In addition to the uncertainty of statistical estimates, there is uncertainty extrapolating from populations to individuals *(42)*. An approach suggested to convey this kind of uncertainty *(43)* is to use qualifying statements, such as:

> "There is no way of knowing whether you will be one of the women who gets breast cancer. In addition, if you take tamoxifen, there is no way of knowing whether you will be one of the women who benefited from it."

An explicit acknowledgment of uncertainty should also accompany messages based on the results of a single study, intermediate end point, or extrapolations across populations. It is rare that a single study provides a definitive answer about a particular question. Consequently, it is critical to put the results of a single study into the context of similar studies and to grade the quality of the evidence (i.e., give less weight to the results of observational studies or subgroup analyses and more weight to randomized trials). Communicators should be particularly cautious about drawing strong inferences when small differences are reported by studies with weaker designs. Intermediate end points should also raise caution because changes in such end points (e.g., tumor shrinkage) may not translate into clinically important improvements in patient outcomes (e.g., improved length or quality of life). An additional concern is whether the study results are really generalizable to the patients the clinic sees: Would this population have met entry criteria for the trial? Is this disease a relatively minor competing risk for this population?

Remember Health

One of the main objectives of medical care is to improve the health of the population. Recently, health communicators have begun to increasingly focus on increasing individual's awareness of the disease risks they face and in identifying strategies to modify these risks. In many cases, compliance with recommended risk reduction strategies (e.g., screening for colorectal cancer) has been considered suboptimal. To improve compliance, a number of public health campaigns now actively seek to persuade the public to adopt specific preventive strategies. Many campaigns use scare tactics to promote a particular behavior (e.g., "feeling well is sometimes the first sign of colon cancer—get screened today!" or "you can't see it or feel it, but you may have cancer").

Whereas such persuasive tactics may elicit intended behavioral changes, they have other consequences that may paradoxically worsen a population's sense of health *(44–47)*. First, rather than promoting a sense of health, such tactics may simply increase everyone's sense of vulnerability and anxiety about disease. These messages make it clear that no one is really healthy. For example, telling the story of a 30-year-old woman with breast cancer may garner a lot of attention and motivate some older women to undergo screening mammography, but it will probably also frighten many young women who stand to gain little if anything from mammography. Second, these aggressive tactics may convey a false sense of the magnitude and certainty of the benefits of interventions, engendering unrealistic expectations. Finally, the heightened emphasis on taking personal responsibility for reducing one's risk may lead people diagnosed with disease to blame themselves.

Ironically, the increasing prevalence of persuasive messages coincides with a shift in contemporary medical ethics to a shared decision-making model in which the clinician's role is not to persuade patients to adopt a particular behavior (i.e., use any means necessary to get them to eat that fifth daily fruit) but, where possible, to help patients understand the risks and benefits of the options they face so they can make informed choices between them (e.g., "I understand the pros and cons, and I choose to eat this ice cream"). This model places increasing emphasis on the role of patient preferences and values in medical decision making—because physicians and patients may have different interpretations of well-being.

Whether the evidence of benefit for an intervention is questionable or certain, we believe it is important to consider the likely net effect of such tactics on well-being. We argue that the fundamental purpose of risk communication is to provide individuals with the facts they need to make informed decisions. Increasing the public's sense of vulnerability to inspire a healthy behavior undermines well-being and may result in net harm. We encourage communicators to be sensitive about the potential side effects of their messages.

Conclusions

Although clinicians clearly need to be part of the solution, competing demands and inadequate training in how to talk to patients about risk limit what clinicians can contribute toward improving the state of medical risk communication. Moreover, use of criteria such as Health Plan Employer Data and Information Set (HEDIS) report cards to measure "the quality of care" by the degree with which practice complies with guidelines rather than on some measure of the quality of decision making creates a perverse incentive to prescribe rather than to discuss treatment options.

By acknowledging the realities of clinical practice, we advocate a three-part plan to improve clinical risk communication.

1) Help clinicians to do more by providing clinicians with simple and efficient office-based tools to generate and display up-to-date risk and benefit estimates

2) Educate patients with a reader's guide for patients to help the public more critically evaluate the ubiquitous health risk data to which they are exposed

3) Disseminate *Guidance for Communicators,* a writer's guide to risk communication to help journalists and public health agencies express risk information in a clear, balanced, and understandable way

Like any intervention, ours will need careful study to evaluate whether it is effective and acceptable to clinicians, communicators, and patients.

Appendix Table 1. Summary or prior randomized trials of disease-specific educational interventions that presented quantitative data and tested comprehension

Health issue (reference No.)	Population	Intervention	Knowledge assessment*	Finding	
Hormone replacement for postmenopausal women *(28)*	165 women recruited for trial from health practitioners, clinics, community advertisement	Intervention: Tailored decision aid Control: ACP education pamphlet from the American College of Physicians	*How many women out of ____ will get "disease" in their lifetime*	% Women with realistic perceived lifetime risk†	
				Decision Aid	ACP Pamphlet
			Risk without HRT		
			CHD	57%	61%
			Hip fracture	81%	74%
			Breast cancer	71%	44%
			Risk with HRT		
			CHD	75%	23%
			Hip fracture	79%	33%
			Breast cancer	79%	51%
Breast cancer risk counseling for first-degree relatives of women with breast cancer *(29)*	200 women identified by relatives under treatment for breast cancer at two major cancer centers	Intervention: 1.5-hour counseling session by trained nurse-educator who told women: "your personal chances of developing breast cancer by age 80 are ___%, or 1 in ___." Control: General health counseling	*Rate your chances of getting breast cancer during your lifetime on a scale from 0 (definitely will not get it) to 100 (definitely will get it)*	% Women extremely overestimating their breast cancer risk‡	
				Intervention	Control
			Before	64%	67%
			After	63%	64%
Prostate cancer screening *(30)*	196 men scheduled for general medical clinic in academic center	Intervention: Shared decision-making video Control: No video	*How many untreated men with early stage prostate cancer would die of this disease?*	% Correct answer	
				Intervention	Control
			Most or all will About half will Most will not	93%	41%
			Does active treatment extend life Very or pretty sure it can Not sure Very or pretty sure it cannot	67%	24%
			How man men with elevated PSA have prostate cancer? Most or all do About half do Most do not	72%	15%
Consent for anesthesia *(31)*	40 patients scheduled to undergo surgery requiring general anesthesia	Intervention: Explicit discussion numerical data Control: Routine care	*Perceived risk elicited with visual analog scale*	% Correct answer	
				Intervention	Control
			Nausea	40%	35%
			Sore throat	75%	75%
			Death	55%	15%
			Brain damage	80%	65%
			Tooth damage	70%	30%
			Inadequate sedation	40%	50%
Correct inaccurate perceptions of the chance of four different health risks: heart attack, stroke, cancer, and motor vehicle accident *(32)*	1317 adult patients recruited from the waiting room of eight family practices	Intervention: Patient estimates perceived risks and is then given computer-generated, individualized feedback derived from population statistics Control: No feedback	*Compared with others your same age and sex, how would you rate your risk of [event] in the next 10 years?* *Using a 5-point scale ranging from "much lower" to "much higher than average"*	Increasing perceived stroke risk among patients who had underestimated Reducing perceived risk of cancer among those who had overestimated No change in perceived heart attack or motor vehicle accident risk	

*CHD = coronary heart disease; HRT = hormone replacement therapy; PSA = prostate-specific antigen.

†"Realistic" means a woman's perceived risk estimate fell in the same quartile as her predicted risk calculated for women with a similar characteristics.

‡Accuracy of risk estimation was calculated as the difference between an individual's subjective estimate and an objective estimate based on the Gail algorithm. "Extreme overestimation" means the subject's estimate exceeded the highest possible lifetime Gail score for a woman of the same age. At follow-up, 89% of control subjects and 82% of intervention patients overestimated their personal risk by at least 10%. Reprinted with permission from Effective Clinical Practice.

References

(1) Veterans Health Administration. Journey of change. Washington (DC): Department of Veterans Affairs; 1997

(2) National Institutes of Health Consensus Development Conference. Breast cancer screening for women ages 40–49: program and abstracts. Bethesda (MD): National Cancer Institute; 1997.

(3) Emanuel EJ, Emanuel LL. Four models of the physician–patient relationship. JAMA 1992;267:2221–6.

(4) National Center for Health Statistics. National ambulatory medical care survey: 1996. Hyattsville (MD): U.S. Department of Health and Human Services; 1997. Available online from: http://www.cdc.gov/nchswwv/data/ad295.pdf

(5) Kalet A, Roberts JC, Fletcher R. How do physicians talk with their patients about risks? J Gen Intern Med 1994;9:402–4.

(6) Hall JA, Roter DL, Katz NR. Meta-analysis of correlates of provider behavior in medical encounters. Med Care 1988;26:657–75.

(7) Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB, et al. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837–47.

(8) O'Connor GT, Plume SK, Olmstead EM, Coffin LH, Morton JR, Maloney CT, et al. Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. Northern New England Cardiovascular Disease Study Group. Circulation 1992;85:2110–8.

(9) National Cancer Institute. Breast cancer risk tool: an interactive patient education tool [software]. Bethesda (MD): National Cancer Institute; 1998.

(10) Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989; 24:1879–86.

(11) Kalichman SC, Coley B. Context framing to enhance HIV-antibody-testing messages targeted to African American women. Health Psychol 1995;14: 247–54.

(12) Skinner CS, Strecher VJ, Hospers H. Physicians' recommendations for mammography: do tailored messages make a difference? Am J Public Health 1994;84:43–9.

(13) Rimer BK, Glassman B. Tailoring communications for primary care settings. Methods Inf Med 1998;37:171–7.

(14) Ries LAG, Kosary CL, Hankey BF, Miller BA, Edwards BK, editors. SEER cancer statistics review, 1973–1996. Bethesda (MD): National Cancer Institute; 1999. Available online from: http://www.seer.ims.nci.nih.gov/

(15) National Center for Health Statistics. Mortality data [online] 1999. Available from: http://www.cdc.gov/nchswww/datawh/statab/upubd/mortabs.htm

(16) Gigerenzer G. The psychology of good judgment: frequency formats and simple algorithms. Med Decis Making 1996;16:273–80.

(17) Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. Acad Med 1998;73:538–40.

(18) Kirsch IS, Jungeblut A, Jenkins L, Kolstad A. Adult literacy in America. Washington (DC): Office of Education Research and Improvement, U.S. Department of Education; 1993.

(19) Schwartz LM, Woloshin S, Black W, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med 1997;127:966–72.

(20) Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM. The framing effect of relative and absolute risk. J Gen Intern Med 1993;8: 543–8.

(21) Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. Am J Med 1992;92: 121–4.

(22) Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med 1992;117:916–21.

(23) Hux JE, Naylor CD. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? Med Decis Making 1995;15:152–7.

(24) O'Connor AM. Effects of framing and level of probability on patients' preferences for cancer chemotherapy. J Clin Epidemiol 1989;42:119–26.

(25) Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981;211:453–8.

(26) McNeil BJ, Pauker SG, Sox HC Jr, Tversky A. On the elicitation of preferences for alternative therapies. N Engl J Med 1982;306:1259–62.

(27) Woloshin S, Schwartz LM. How can we help people make sense of medical data? Eff Clin Pract 1999;2:176–83.

(28) O'Connor AM, Tugwell P, Wells GA, Elmslie T, Jolly E, Hollingworth G, et al. Randomized trial of a portable, self-administered decision aid for postmenopausal women considering long-term preventive hormone therapy. Med Decis Making 1998;18:295–303.

(29) Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995;87:286–92.

(30) Flood AB, Wennberg JE, Nease RF Jr, Fowler FJ Jr, Ding J, Hynes LM. The importance of patient preference in the decision to screen for prostate cancer. J Gen Intern Med 1996;11:342–9.

(31) Inglis S, Farnill D. The effects of providing preoperative statistical anaesthetic-risk information. Anaesth Intensive Care 1993;21:799–805.

(32) Krueter MW, Strecher VJ. Changing inaccurate perceptions of health risk: results from a randomized trial. Health Psychol 1995;14:56–63.

(33) Halpern DF, Blackman S, Salzman B. Using statistical risk information to assess oral contraceptive safety. App Cogn Psychol 1989;3:251–60.

(34) Black WC, Welch HG. Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. N Engl J Med 1993;328:1237–43.

(35) Black W. Should this patient be screened for cancer? Eff Clin Pract 1999; 2:86–95.

(36) Freimuth VS, Greenberg RH, DeWitt J, Romano R. Covering cancer: newspapers and the public interest. J Communication 1983;34:62–73.

(37) Oxman AD, Guyatt GH, Cook DH, Jaeschke R, Heddle N, Keller J. An index of scientific quality for health reports in the lay press. J Clin Epidemiol 1993;46:987–1001.

(38) Slaytor EK, Ward JE. How risks of breast cancer and benefits of screening are communicated to women: analysis of 58 pamphlets. BMJ 1998;317: 263–4.

(39) Woloshin S, Schwartz LM, Byram SJ, Fischhoff B, Welch HG. Scales for assessing perceptions of event probability: a validation study. Med Decis Making 1998;14:490.

(40) Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavahah M, Cronin WM et al. Tamoxifen for prevention of breast cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst 1998;90:1371–88.

(41) Lichtenstein S, Fischhoff B, Phillips LD. Calibration of probabilities: the state of the art to 1980. In: Kahneman D, Slovic P, Tversky A, editors. Judgment under uncertainty: heuristics and biases. Cambridge (UK): Cambridge University Press; 1982. p. 306–34.

(42) Steiner JF. Talking about treatment: the language of populations and the language of individuals. Ann Intern Med 1999;130:618–22.

(43) Skolbekken JA. Communicating the risk reduction achieved by cholesterol reducing drugs. BMJ 1998;316:1956–8.

(44) Feldman W. How serious are the adverse effects of screening? J Gen Intern Med 1990;5(5 Suppl):S50–3.

(45) Welch HG, Burke W. Uncertainties in genetic testing for chronic disease. JAMA 1998;280:1525–7.

(46) Fitzgerald FT. The tyranny of health. N Engl J Med 1994;331:196–8.

(47) Meador CK. The last well person. N Engl J Med 1994;330:440–1.

Notes

[1]SEER is a set of geographically defined, population-based, central cancer registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis, and the NCI makes the data available to the public for scientific research.

Supported by U.S. Army Medical Research and Materiel Command Breast Cancer Research Program New Investigator Award DAMD17-96-MM-6712 as well as by VA Career Development Awards in Health Services Research and Development.

The views expressed herein do not necessarily represent the views of the Department of Veterans Affairs or the United States Government.

Drs. Schwartz and Woloshin are joint first authors of this manuscript. The order of their names is entirely arbitrary.

Interactive Multimedia and Risk Communication

Victor J. Strecher, Todd Greenwood, Catharine Wang, Dana Dumont

As our understanding of risk factors and their interaction with individual susceptibility to disease improves, general messages designed to communicate risk seem increasingly ineffective and often misleading. Risk messages communicated through the mass media cannot convey an individual's personal susceptibility to preventable diseases or the seriousness of these diseases. The advent of new media technologies allows us to better reach the public with programs tailored to the needs and interests of individual users. Although similar in outward appearance to mass media, programs delivered through the Internet, CD-ROM, and computer kiosks offer the potential for vastly improved efficacy in communicating risk. This paper outlines the potential uses of interactive multimedia within the traditional goals of risk communication. A significant research endeavor, coupled with stronger avenues for dissemination, is recommended to achieve the potential of new media in a timely manner. [Monogr Natl Cancer Inst 1999;25:134–9]

Introduction

The concept of perceived risk is central to most models of health behavior change and decision making *(1–4)*. Public health campaigns designed to improve health-related behaviors or to stimulate awareness of a health issue at a population level often employ risk-based messages. However, disseminating risk messages to the general public in ways that are accurate, clear, and attention grabbing is a challenging task *(5)*. Bottorff et al. *(6)* found that, although population-based information about risk factors for cancer is available, there are few opportunities for individuals to obtain accurate information that is personalized and based on individualized risk factors. Without knowledge of an individual's actual risk, risk-based communication is easily disregarded by those at high risk, while taken too seriously by those at low risk.

As our understanding of risk factors and their interaction with individual susceptibility to disease improves, general messages designed to communicate risk seem increasingly ineffective and often misleading. The general public, by now, is inundated with the "risk factor of the week" and has grown weary of our attempts to generate media hype for these factors *(7,8)*. The current "broadband" influx of health risk information may be likened to filling a water glass with a fire hose. Individuals may feel overwhelmed by the sheer number of "bad things" they need to improve. Mason et al. *(9)* state that "the very multiplicity of threats and the urgency with which they are presented make it difficult for most of us to sort out major from minor, proven from suspected, and most importantly, those that we as individuals can control from those we cannot." Mass media fail to effectively convey the personal susceptibility to preventable diseases or the seriousness of these diseases; they provide messages that are easy to ignore *(10)*. Often, the segments of the public who are the least informed about potential hazards are the very people who are likely to selectively ignore messages aimed at them *(11)*.

Interactive multimedia[1] have the potential to assist users in sorting out major from minor, proven from suspected, and controllable from uncontrollable risks. These new media systems allow us to reach the public more effectively with programs tailored to the needs and interests of individual users. Whereas the outward appearances of interactive multimedia are similar to those of mass media, interactive programs delivered through the Internet, CD-ROM, or computer kiosks offer the potential for vastly improved efficacy in communicating risk. Evaluations of computer-based health education programs in medical and educational settings suggest that these programs can be effective in changing behavioral intentions, influencing health behaviors, and improving the management of disease *(13–16)*.

The purpose of this paper is to outline the potential uses of interactive multimedia within the traditional goals of risk communication. We hope that the discussion will stimulate new research and higher quality programs for the public. We begin with a discussion of the general advantages of interactive multimedia, followed by the potential for interactive multimedia in (*a*) presenting risk information in an accurate, understandable manner; (*b*) reducing inaccurate perceptions of personal risk; and (*c*) facilitating better informed health decisions. We close with a discussion of methods for reaching the public with these important programs.

Why Use Interactive Multimedia?

There are several advantages to using interactive multimedia. Three notable advantages—interactivity, adaptiveness, and assessment—are relevant to the communication of risk information.

Interactivity

The systems we are considering in this paper require active, versus passive, involvement of the user. Active involvement with a message has been associated with a higher degree of active information processing *(17)*. In one study of classroom use of educational technology *(18)*, students who used computer-based instruction, in contrast to those who used more traditional curricular materials (i.e., textbooks), tended to have a higher percentage of "time on task" and were generally more involved in the process of learning.

It is likely that individuals vary in their preferences for sequencing, pacing, feedback, and review of interactive multimedia messages, although far more research is required in this area. Preferences for a rapidly paced multimedia experience among

Affiliation of authors: University of Michigan Comprehensive Cancer Center, Health Media Research Laboratory, Ann Arbor.

Correspondence to: Victor J. Strecher, Ph.D., M.P.H., Health Media Research Laboratory, 300 N. Ingalls, Suite 5D, Ann Arbor, MI 48109-0471.

See "Note" following "References."

© Oxford University Press

adolescent boys can readily be observed in any video arcade. The experience of our Health Media Research Laboratory with focus groups of middle school adolescents and of older adults using interactive multimedia programs suggests that adolescents prefer fast-paced interactivity than do the older adults. Tests of other sociodemographic groups, as well as the influence of pacing on relevant outcomes, should be conducted in the health area.

Adaptiveness

Another advantage of interactive multimedia is the ability to adapt, or tailor, information to relative characteristics of the user. By simply asking questions, interactive multimedia programs can generate an educational experience tailored to the user's social and cultural milieu, sociodemographic characteristics, relevant life events, literacy level, concept recognition abilities and preferences, risk perceptions, physical and functional health status, health-related behaviors, health knowledge, perceived barriers to and benefits of health behavior changes, stage of health behavior change, self-efficacy, and attributions for previous failures, among a universe of possible variables. Never has the understanding and application of theories of individual health behavior been so relevant; we now have the potential to consistently apply these theories and their related concepts to improve the health of the public through individualized messages.

Assessment

Interactive multimedia programs are uniquely suited for comprehensive assessment. First, their interactivity allows significant shortening in the length of most questionnaire instruments. (If, on a health risk assessment, we find that the user does not eat meat, he or she no longer needs to be asked questions about meat.) While skip patterns can be employed in print-based questionnaires, computer-based questionnaires are less prone to respondent error *(19)*. Interactive multimedia have the advantage of seamlessly constructing complex hierarchies, invisible to the user, that significantly shorten assessment forms. Second, the perceived anonymity of computer-based assessment programs has been shown to yield dramatic improvements in the quality of survey measurements, especially for surveys of sexual, drug, and other sensitive behaviors *(20)*. Accurate estimation of epidemiologic risk, especially for sensitive behaviors, has been limited in the past by the need to report socially undesirable risk factors to others *(21)*. When computers ask questions, respondents have reported that they believe their information will be kept confidential *(22)*. Third, interactive multimedia programs can determine progress made. Most multimedia authoring tools readily support the collection, storage, and processing of all interactions made with an instructional computer program. Preintervention and postintervention assessments can therefore be created as a seamless part of the program itself.

Application to Traditional Goals of Risk Communication

Presenting Risk Information

Johnson and Slovic *(23)* found that, although people are unfamiliar with the concept of uncertainty, they are more likely to recognize uncertainty when it is presented simply and when meaningful graphics are used. In addition, uncertainty and probability may be difficult for the lay person to understand because they include the element of time. Genetic counselors, for example, are often faced with the task of explaining cumulative probabilities to patients. With static media, genetic counselors are limited in their ability to demonstrate how a patient's risk for disease changes as he or she ages *(24)*. Multimedia presentations may enhance a counselor's ability to communicate this information to patients through the use of graphic animations that can allow clients to explore how risk and time interact. Animation can reinforce the concept that change in one's estimated risk for disease over time is a gradually changing estimate. The interactive CD-ROM called "Counseling by Computer: Breast Cancer Risk and Genetic Testing" provides such an animation, showing how risk for cancer changes throughout life *(25)*. In addition, this program also includes a valuable education component on the meaning of risk estimates by incorporating lessons using animated coin tosses to demonstrate chance and probability.

Risk ladders or other visual devices emphasize how the risk for one hazard compares with risks for other more commonly understood hazards *(26,27)*. Although there are many pitfalls in comparing one hazard to another, even when it occurs at the same frequency risk [*see (5)*, page 96, for discussion], interactive multimedia can present comparisons in a way that is less misleading. Features, such as animation, zooming in and out on graphics, and the provision of hyperlinks to meaningful explanations about the limitations of comparable data *(28)*, can help alleviate some of the more common problems with risk comparisons. Multimedia can deliver integrated messages about lowering risk and the effect of actions on risk estimations. Interactive computer programs are able to compare estimated risk under current circumstances with potential future conditions. This instant "what if" comparison cannot be easily made when delivering risk information using other channels, including human interaction. "What if" comparisons are useful in presenting risk information aimed either at altering risky behaviors or at providing information for a decision under conditions of uncertainty *(29)*.

Reducing Inaccurate Perceptions of Risk

Risk is often communicated for the purpose of educating individuals who are either unaware that they are at risk for a condition or who believe that they are at lower risk than their epidemiologic risk would indicate. Individuals frequently display an "optimistic bias" of their risk of being affected by a hazard *(30,31)*. For example, Strecher et al. *(32)* found that, although most smokers accurately perceived their risks to be greater than nonsmokers for smoking-related illness, they were more likely to underestimate their risk and thus misunderstand the magnitude of smoking as a risk factor for disease. Related to unrealistic optimistic bias is overconfidence in one's judgment of risk. Individuals are often sure that they "know" the odds; overconfidence reduces the perceived need for and the perceived importance of learning about risk *(33)*.

Kreuter and Strecher *(34)* demonstrated that inaccurate perceptions of risk could be modified toward a more realistic view using computer-tailored printed messages by comparing their perceived risk with their actual estimated risk. Comparisons of perceived versus actual risk have been used successfully in multimedia as well. "Straight Talk About Breast Cancer" is a program designed to foster realistic perception of breast cancer risk among women *(35)*. The program assesses women's perceived risk and then calculates an immediate assessment of actual risk,

using the risk estimation model of Benichou et al. *(36)*. These actual and perceived risk estimations are then graphically compared. If they are not similar, an explanation of how actual risk was determined is presented. A pilot study, using a randomized trial design to test the effect of "Straight Talk About Breast Cancer" found that 50% of women who initially had an optimistic bias of their breast cancer risk at the time of the intervention were realistic at follow-up, that 55% of women who had a pessimistic bias of their breast cancer risk at intervention reported a realistic bias at follow-up, and that 80% of those women who were realists at the time of intervention were still realists at follow-up *(35)*. Altering pessimistic bias is also an important goal of interactive multimedia, because having a pessimistic bias of one's risk may result in overuse of health care services *(34,37)*.

Risk perceptions may be more likely to change if the message is presented in a vivid manner. Nisbett and Ross *(38)* state that vivid information is (*a*) emotionally interesting; (*b*) concrete and imagery provoking; and (*c*) proximate in a sensory, temporal, or spatial way. Multimedia programs achieve these characteristics of vividness, allowing the user to see graphic images of disease severity or to hear testimonials of individuals who have experienced a particular disease. "The Michigan Interactive Health Kiosk Project," a network of 100 public-access multimedia health information stations, provides the general public with an opportunity to learn about actual risk for breast cancer, prostate cancer, and heart disease. These estimates, based on age, family history, and other risk factors, are provided to the user in a matter of seconds. Our "Michigan Interactive Health Kiosk Project" includes a "Scare Me" button in the cigarette smoking section of the kiosk. This button is pressed more than any other button of the section. Other graphic depictions of disease severity are used in the sexually transmitted disease (STD) section and in the childhood immunization section. In each of these sections, threat-reducing (e.g., "How Do I Quit" options) programming is also included.

In an analysis of the research literature, Taylor and Thompson *(39)* found that case histories tend be more persuasive than similar information presented as statistics or straight facts. Brosius and Bathelt *(40)* looked at the relationship of exemplars to base-rate risk information and found that the number and quality of exemplars had a strong effect on one's view of the importance of a problem. Interactive multimedia clearly have potential to present information through ease histories and testimonials; in fact, these case histories may be tailored to the specific background, interests, and style of the user. Noell et al. *(41)* used personal scenarios in an interactive videodisc designed to develop decision-making skills for adolescents on preventing STDs. Use of the prevention program increased risk perception associated with unprotected sex.

Facilitating Informed Decision Making

A primary objective of risk communication is to provide individuals with relevant information, enabling them to make informed medical decisions. The goal of risk communication, in this case, is not to persuade an individual to make a specific choice but to provide risk information in a "neutral" or nondirective manner so that the individual can then decide what is best. Typically, this form of risk communication occurs when an individual is faced with an important medical decision (e.g., whether to have hormone therapy) in which the obvious choice is not clear and satisfaction with the outcome of the decision depends on the values held by the patient.

Interactive multimedia programs can allow individuals to weigh their attitudes toward potential outcomes, informed by risk and their symptoms. One example of this approach is the "Shared Decision Making Program on Benign Prostatic Hyperplasia (BPH)" *(42)*. This multimedia decision-making program shows personal case stories and probability information relevant to various BPH outcomes that are tailored to the individual. Risk is tailored on the basis of age, general health status, symptom level, and history of acute retention. Patients in this study were asked to rank the degree to which they were bothered by their symptoms and their attitudes about impotence. The relative importance of these factors was measured against the objective level of the patients' actual symptoms in the decision to choose surgery or watchful waiting. The result was that symptom levels no longer predicted choice of treatment; instead, the shared decision program helped patients make a choice that reflected their attitudes about their symptoms and their concern about risks *(42)*. "The Shared Decision Making Program on BPH" also uses interviews with patients who have had complications: one interview with a patient who chose watchful waiting and one who chose surgery. Wennberg states [*(42);* page 118] that, by using an integrated decision-making tool such as "Shared Decision Making Program on BPH," "patients become empowered to choose their own treatments."

Tremendous opportunities for interactive multimedia programs exist in the area of informed decision making. Interactive multimedia programs have also been created for patients undergoing treatment for breast cancer and for human immunodeficiency virus *(43,44)*. These programs include decision aids that help individuals clarify values, learn about options, and examine the consequences of their decisions. In addition, Internet-based programs have been created to assist individuals in making decisions about health care plans and professionals *(45,46)*. In a broad array of illnesses and preventive health behaviors, complex decisions must be made on the basis of an analysis of medical and epidemiologic data linked to personal values and preferences. For example, in genetic screening and counseling, numerous informed decisions must be made regarding decisions to undergo genetic testing and other related medical decisions, such as prophylactic surgery *(47,48)*. These decisions are no longer left simply to health providers or public health professionals; health consumers are likely to increasingly demand a stronger role in decision making accompanied by greater access to relevant health information *(49–51)*.

Reaching the Public With Interactive Multimedia

Consumers' demand for health information is steadily increasing. McGinnis et al. *(52)* report that telephone inquiries to the Public Health Service's health information clearinghouses more than doubled in the early 1990s and that mail inquiries grew by 43%. More than two thirds of consumers have questions about their personal health *(52)*, yet a 1994 survey published by the Medical Library Association *(53)* found that nearly 70% of respondents reported problems in gaining access to relevant health information.

It is likely that, within the next 5–10 years, traditional mass communication channels, such as television and print media, will evolve into interactive television channels, creating oppor-

tunities to disseminate health programs that personalize messages to specific risk factors, stages of change, and health beliefs of users *(54)*. In the June 1995 issue of the *American Journal of Public Health,* C. Everett Koop states that "cutting-edge technology, especially in communication and information transfer, will enable the greatest advances yet in public health. . . . Eventually, personal home tele-medicine links could provide every home with access to health information 24 hours a day, 7 days a week, encouraging personal wellness and prevention and leading to better informed decisions about health care. A generation of children raised on video games will probably be more attuned to health messages coming from interactive videos than from lectures by the school nurse" *(55)*.

The bridge to this future is the current interactive multimedia environment—increasingly through the Internet. According to a recent survey *(56)* conducted jointly by the Institute for the Future in Menlo Park, CA, and Princeton Survey Research Associates, nearly two thirds of the people who use the World Wide Web seek medical information and referrals. According to another survey *(57)* from the Pew Research Center for the People and the Press, in the absence of a big breaking story, 64% of those surveyed suggested that they go to the Internet to find news on science and health.

In the fall of 1995, the U.S. General Accounting Office conducted a survey *(58)* of 80 experts in the area of consumer health informatics to determine key issues as our society further develops advanced communications technologies in the consumer health area. This survey identified a number of significant barriers, requiring attention as consumer health informatics continues to develop. The most commonly cited barrier (reported by 79% of the experts) was lack of access. Access to and use of health information in general are directly tied to socioeconomic status, and these factors are tied to health status. The poor have cancer survival rates 10%–15% less than higher socioeconomic groups. Socioeconomic status is positively related to the practice of preventive health behaviors *(59)*. In general, individuals with lower socioeconomic status know less about cancer prevention, such as early warning signs; they know less about the health care system, are not as aware of prevention services, and are more likely to rely on folk medicine and nonstandard, alternative treatments *(60,61)*. The difference between the information haves and have-nots in the new information economy has brought the concepts of the information poor and the information gap to the forefront as important policy issues *(62,63)*. The Internet creates a more privatized information environment, available only to those who have the resources to access it.

In general, individuals seek out information that is the most accessible *(64)*. Accessibility may be the single most important criterion in evaluation by users of an information system *(65)*, outweighing the perceived quality of the information *(66)* and credibility *(64)*. Research by Allen *(67)* showed that even highly educated professionals will knowingly seek out inferior information from a more accessible source. The field research by Allen found that the physical distance between the individual and the information source predicted usage far more than the perceived quality of the answer.

"The Michigan Interactive Health Kiosk Project" is one example of how interactive multimedia technology can be made available to a broader spectrum of the public. Health 'o' Vision is an interactive multimedia, touchscreen kiosk that is accessible at no cost by any person in the vicinity of the kiosk who is curious about it. There are 100 kiosks currently providing in-depth health information on 10 topics, four of which provide tailored estimation of risk for disease (breast cancer, prostate cancer, coronary heart disease, and STD). There is no process by which people are either preselected to be directed to the information or screened to be excluded from information. Kiosks are located throughout the state of Michigan in urban, suburban, and rural locations. They are placed in high-traffic, publicly accessible locations, such as community centers, supermarkets, shopping malls, YMCAs, and local health departments.

Pilot assessments of users showed that over 400 000 individuals use the kiosks each year. When comparing users with the population exposed to the kiosks, we find that kiosk users tend to be younger (over 50% of users are under 21 years of age). Users do not, however, differ from nonusers by ethnic or gender status. Moreover, individuals who use the kiosk are more likely to smoke cigarettes (approximately 37% of the kiosk users were cigarette smokers) than do nonusers. There is probably a higher proportion of smokers to nonsmokers than that found on the Internet *(68)*, although there is a significant lack of data characterizing the health behaviors of Internet users. Because of the ethnic compositions where the kiosks are placed, over 50% of kiosk users are nonwhite. Satisfaction levels with the kiosks do not differ by ethnic status or by gender. These data suggest that interactive multimedia would be used by the public most in need of preventive services—those who do not have ready access to computers.

Summary

Interactive multimedia systems, such as the Internet, CD-ROM, and kiosk programs, have the potential for reaching populations in need of risk communication messages. But it is clear that we must determine (*a*) how we can harness the power of new media to create the most efficacious programs and (*b*) how we can maximize the exposure, access, and reach of these programs.

The revolutionary advances in our ability to predict, *at the level of the individual,* health problems before they occur must be matched by our ability to effectively communicate risk at this level. The equally revolutionary advances in information technology allow us to do just that—to communicate risk to mass populations at the level of the individual. An area that requires particular research attention is the development of improved statistical models for the estimation of risk [e.g., *(69)*]. These models should include algorithms for estimating the interactions of multiple risk factors *(70)*.

With increasingly powerful risk assessment tools, interactive multimedia technologies should address the difficulties many end users have in accurately comprehending and recalling risk factor information *(71)*. Interactive multimedia allow us to develop programs tailored to the relevant learning styles of the user, including variations in pacing, in sequencing, and in type of feedback. This technology allows us to generate standardized methods for reviewing previously presented information, additional instruction, and skill building. The graphic capabilities of interactive multimedia allow greater vividness of risk-related questions and messages through digital video, animation, sound, and text.

In the future, interactive multimedia technologies will likely reach many individuals who would not normally purchase or use a computer. However, before the time that interactive multime-

dia channels become as ubiquitous as television, video cassette recorders, and telephones, significant effort will be required to make tailored and interactive health materials available to a broader spectrum of our society. Equitable access will require not only available technology but also content that is understandable, interesting, and relevant.

The degree to which we capitalize on the potential of interactive multimedia in communicating risk will depend largely on the quantity and quality of research in this area. To date, there are very few research groups dedicated to pursuing the many complex questions, challenges, and opportunities posed by the information revolution *(72,73)*. Lack of significant investment in a technology infrastructure and the integrated involvement of health behavior, epidemiologic, technical, and communications expertise has impeded rapid development of this research field. Although front-end costs to create this infrastructure are high, the resulting interventions can be inexpensively disseminated and broadly adopted at low cost. Heightened prioritization from federal and private research agencies is required to develop and to support this nascent field. Such support is likely to result in a payoff equal to or greater than any public health communications effort to date.

References

(1) Becker MH. The health belief model and personal health behavior. Health Educ Monogr 1974;2:324–508.

(2) Eraker SA, Becker MH, Strecher VJ, Kirscht JP. Smoking behavior, cessation techniques, and the health decision model. Am J Med 1985;78: 817–25.

(3) Fishbein M, Ajzen I. Belief, attitude, attention, and behavior: an introduction to theory and research. Reading (MA): Addison-Wesley; 1975.

(4) Rogers RW. Cognitive and physiological processes in fear appeals and attitude change: a revised theory of protective motivation. In: Cacioppo J, Petty, R, editors. Social psychophysiology: a sourcebook. New York (NY): Guilford Press; 1983. p. 153–76.

(5) National Research Council. Improving risk communication. Washington (DC): National Academy Press; 1989.

(6) Bottorff JL, Ratner PA, Johnson JL, Lovato CY, Joab SA. Communicating cancer risk information: the challenges of uncertainty. Pat Educ Couns 1998;33:67–81.

(7) Becker, MH. The tyranny of health promotion. Public Health Rev 1986; 14:15–23.

(8) Covello VT, von Winterfeldt D, Slovic P. Communicating scientific information about health environmental risks: problems and opportunities from a social and behavioral perspective. In: Covello V, Moghissi A, Uppuluri VRR, editors. Uncertainties in risk assessment and risk management. New York (NY): Plenum Press; 1993.

(9) Mason JO, Ogden HG, Berreth DA, Martin LY. Interpreting risks to the public. Am J Prev Med 1986;2:133–9.

(10) Pechmann C, Stewart DW. The multidimensionality of persuasive communication: theoretical and empirical foundations. In: Cafferata P, Tybout AM, editors. Cognitive and affective responses to advertising. Lexington (MA): Lexington Books; 1989. p. 31–56.

(11) Atkin C, Arkin EB. Issues and initiatives in communicating health information to the public. In: Atkin C, Wallack L, editors. Mass communication and public health: complexities and conflicts. Newbury Park (CA): Sage; 1990. p. 7–12.

(12) Galbreath J. The educational buzzword of the 1990s: multimedia, or is it hypermedia, or interactive multimedia, or . . .? Educ Technol 1992;32: 15–9.

(13) Glasgow RE, La Chance PA, Toobert DJ, Brown J, Hampson SE, Riddle MC. Long-term effects and costs of brief behavioural dietary intervention for patients with diabetes delivered from the medical office. Pat Educ Couns 1997;32:175–84.

(14) Rubin DH, Leventhal JM, Sadock RT, Letovosky E, Schottland P, Clemente I, et al. Educational intervention by computer in childhood asthma: a randomized clinical trial testing the use of a new teaching intervention in childhood asthma. Pediatrics 1986;77:1–10.

(15) Robinson T. Community health behavior change through computer network health promotion: preliminary findings from Stanford Health-Net. Comput Methods Programs Biomed 1989;30:137–44.

(16) French D. Influence smoking cessation with computer-assisted instruction. AAOHN J. 1986;34:391–4.

(17) Petty RE, Cacioppo JT. Communication and persuasion: central and peripheral routes to attitude change. New York (NY): Springer-Verlag; 1986.

(18) Hasselbring T, Sherwood B, Bransford J, Fleenor K, Griffith D, Goin L. An evaluation of a level-one instructional videodisc program. J Educ Tech Systems 1987;16:151–69.

(19) Kovera CA, Anger WK, Campbell KA, Binder LM Storzbach D, Davis KL, et al. Computer-administration of questionnaires: a health screening system (HSS) developed for veterans. Neurotoxicol Teratol 1996;18:511–8.

(20) Cooley PC, Turner CF. Implementing audio-CASI on Windows platforms. Comput Hum Behav 1998;14:195–207.

(21) Renn O, Levine D. Credibility and trust in risk communication. In: Kasperson R, Stallen P, editors. Communicating risks to the public: international perspectives. Dordrecht (The Netherlands): Kluwer Academic Publishers; 1991. p. 175–218.

(22) Shakeshaft AP, Bowman JA, Sanson-Fisher RW. Computers in community-based drug and alcohol clinical settings: are they acceptable to respondents? Drug Alcohol Depend 1998;50:177–80.

(23) Johnson BB, Slovic P. Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. Risk Anal 1995; 15:485–94.

(24) Baty BJ, Venne VL, McDonald J, Croyle R, Halls C, Nash JE, et al. BRCA1 testing: genetic counseling protocol development and counseling issues. J Gen 1997;6:223–44.

(25) Green MJ, Fost N. An interactive computer program for educating and counseling patients about genetic susceptibility to breast cancer. J Cancer Educ 1997;12:204–8.

(26) Roth E, Morgan MG, Fischhoff B, Lave L, Bostrom A. What do we know about making risk comparisons? Risk Anal 1990;10:375–87.

(27) Sandman PM, Weinstein ND, Miller P. High risk or low: how location on a "risk ladder" affects perceived risk. Risk Anal 1994;14:35–45.

(28) Jonassen DH. Hypertext principles for text and courseware design. Educ Psychol 1986;21:269–92.

(29) Janis IL. The patient as decision maker. In: Gentry WD, editor. The handbook of behavioral medicine. New York (NY): Guilford; 1984. p. 326–68.

(30) Weinstein ND. Why it won't happen to me: perceptions of risk factors and susceptibility. Health Psychol 1984;3:431–57.

(31) Kulik JA, Mahler HI. Health status, perceptions of risk, and prevention interest for health and nonhealth problems. Health Psychol 1987;6:15–27.

(32) Strecher VJ, Kreuter MW, Kobrin SC. Do cigarette smokers have unrealistic perceptions of their heart attack, cancer, and stroke risks? J Behav Med 1995;18:45–54.

(33) Slovic P, Fischhoff B, Lichtenstein S. Facts and fears: understanding perceived risk. In: Schwing RC, Albers WA Jr, editors. Societal risk assessment: how safe is safe enough? New York (NY): Plenum; 1980. p. 181–216.

(34) Kreuter MW, Strecher VJ. Changing inaccurate perceptions of health risk: results from a randomized trial. Health Psychol 1995;14:56–63.

(35) Crawford YS, Bulger DW, Kobrin SC, Ciloglu GA, Strecher VJ. Can computer-tailored messages guide women toward more realistic breast cancer risk perception? Paper presented at the 122nd annual meeting of the American Public Health Association; 1994 Oct 30–Nov 3; Washington (DC).

(36) Benichou J, Gail MH, Mulvihill JJ. Graphs to estimate an individualized risk of breast cancer. J Clin Oncol 1996;14:103–10.

(37) Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995;87:286–92.

(38) Nisbett RE, Ross L. Human inference: strategies and shortcomings of social judgment. Englewood Cliffs (NJ): Prentice-Hall; 1980.

(39) Taylor S, Thompson S. Stalking the elusive "vividness" effect. Psychol Rev 1982;89:155–81.

(40) Brosius H, Bathelt A. The utility of exemplars in persuasive communications. Comm Res 1994;21:48–78.

(41) Noell J, Ary D, Duncan T. Development and evaluation of a sexual decision-making and social skills program: "the choice is yours—preventing HIV/STDs". Health Educ Behav 1997;24:87–101.
(42) Wennberg J. Shared decision making and multimedia. In: Harris LM, editor. Health and the new media: technologies transforming personal and public health. Mahwah (NJ): Lawrence Erlbaum; 1995. p. 109–26.
(43) Bosworth LK, Gustafson DH. CHESS: providing decision support for reducing health risk behavior and improving access to health services. Interfaces 1991;21:93–104.
(44) Gustafson DH, Hawkins R, Boberg E, Pingree S, Serlin RE, Graziano R, et al. Impact of a patient-centered, computer-based health information/support system. Am J Prev Med 1999; 16:1–9.
(45) Firshein J. US physicians' malpractice data goes on Internet. Lancet 1997; 349:1155.
(46) Meyer H. Information systems: surfing the Net for a health plan. Hosp Health Netw 1996;70:37–8.
(47) Lerman C, Narod S, Schulman K, Hughes C, Gomez-Caminero A, Bonney G, et al. BRCA1 testing in families with hereditary breast–ovarian cancer. A prospective study of patient decision making and outcomes. JAMA 1996; 275:1885–92.
(48) Lerman C, Schwartz MD, Lin TH, Hughes C, Narod S, Lynch HT. The influence of psychological distress on use of genetic testing for cancer risk. J Consult Clin Psychol 1997;65:414–20.
(49) Gustafson DH, Robinson TN, Ansley D, Adler L, Brennan PF. Consumers and evaluation of interactive health communication applications. The Science Panel on Interactive Communication and Health. Am J Prev Med 1999;16:23–9.
(50) Herzlinger RE. Market-driven health care: who wins, who loses in the transformation of America's largest service industry. Reading (MA): Addison-Wesley; 1997.
(51) O'Connor AM, Fiset V, DeGrasse C, Graham ID, Evans W, Stacey D, et al. Decision aids for patients considering options affecting cancer outcomes: evidence of efficacy and policy implications. Mongr Natl Cancer Inst 1999;25:67–80.
(52) McGinnis JM, Deering MJ, Patrick K. Public health information and the new media: a view from the Public Health Service. In: Harris LM, editor. Health and the new media: technologies transforming personal and public health. Mahwah (NJ): Lawrence Erlbaum; 1995. p. 127–41.
(53) Deering MJ, Harris J. Consumer health information demand and delivery: implications for libraries. Bull Med Libr Assoc 1996;84:209–16.
(54) Skinner CS, Siegfried JC, Kegler MC, Strecher VJ. The potential of computers in patient education. Pat Educ Couns 1993;22:27–34.
(55) Koop CE. A personal role in health care reform [editorial]. Am J Public Health 1995;85:759–60.
(56) Hafner K. Can the Internet cure the common cold? New York Times 1998 Jul 9;G1:147.
(57) Barringer F. Media. New York Times 1998 Jul 27;D8:147.
(58) General Accounting Office. Report to the Chairman, Subcommittee on Human Resources and Intergovernmental Relations, House Committee on Government Reform and Oversight. Consumer health informatics: emerging issues. Washington (DC): Government Accounting Office; 1996 Report No.: (GAO/AIMD-96-86).
(59) Nemcek MA. Health beliefs and preventive behavior. A review of the research literature. AAOHN J 1990;38:127–38.
(60) Freimuth VS, Stein JA, Kean TJ. Searching for health information: the Cancer Information Service model. Philadelphia (PA): University of Pennsylvania Press; 1989.
(61) Freimuth VS. The chronically uniformed: closing the knowledge gap in health. In: Ray EB, Donohew L, editors. Communication and health: systems perspective. Hillsdale (NJ): Lawrence Erlbaum Associates; 1990. p. 171–86.
(62) Eng TR, Maxfield A, Patrick K, Deering MJ, Ratzan SC, Gustafson DH. Access to health information and support: a public highway or a private road? JAMA 1998;280:1371–5.
(63) Dervin B. Communication gaps and inequities: moving toward a reconceptualization. In: Dervin B, Voight MJ, editors. Progress in communication sciences. Vol. 3. Norwood (NJ): Ablex; 1980. p. 73–112.
(64) Johnson JD. Cancer-related information seeking. Cresskill (NJ): Hampton Press; 1997.
(65) Rice RE, Shook DE. Relationships of job categories and organizational levels to use of communication channels, including electronic model: a meta-analysis and extension. J Management Studies 1990;27:195–229.
(66) Choo CW, Auster E. Environment scanning: acquisition and use of information by managers. In: Williams ME, editor. Annual review of information science and technology. Vol. 28. Medford (NJ): Learned Information; 1993. p. 279–314.
(67) Allen TJ. Managing the flow of technology: technology transfer and the dissemination of technological information within the R&D organization. Cambridge (MA): MIT Press; 1977.
(68) Egan PA, Duffy L. HRAs online: health-related characteristics of web surfers. Paper presented at the 34th annual meeting of the Society of Prospective Medicine; 1998 Oct 25–28; Newport (RI).
(69) Merrill RM, Kessler LG, Udler JM, Rasband GC, Feuer EJ. Comparison of risk estimates for selected diseases and causes of death. Prev Med 1999; 28:179–93.
(70) Daly MB, Lerman CL, Ross E, Schwartz MD, Sands CB, Masny A. Gail model breast cancer risk components are poor predictors of risk perception and screening behavior. Breast Cancer Res Treat 1996;41:59–70.
(71) Breuer B, Kash KM, Rosenthal G, Diemer K, Osborne MP, Miller DG. Reporting bilaterality status in first-degree relatives with breast cancer: a validity study. Gen Epidemiol 1993;10:245–56.
(72) Robinson TN, Patrick K, Eng TR, Gustafson D. An evidence-based approach to interactive health communication: a challenge to medicine in the information age. Science Panel on Interactive Communication and Health. JAMA 1998;280:1264–9.
(73) Science Panel on Interactive Communication and Health. Wired for health and well-being: the emergence of interactive health communication. In: Eng TR, Gustafson DH, editors. Washington (DC): US Department of Health and Human Services, Gov Print Off, April 1999.

Note

[1]Mulitmedia have been defined and redefined; however, as Galbreath *(12)* states: "whether one is a computer, video, or technology specialist, all seem to view multimedia as the combination of two or more media and imply or assume interactivity via a computer."

Is There a Use for Tailored Print Communications in Cancer Risk Communication?

Barbara K. Rimer, Bernard Glassman

The manner of presentation of cancer risk information is critical to its understanding and acceptance by the individual recipient. Optimal communication of cancer risk information must effectively translate the technical meaning and subtleties of risk and its associated factors to a conceptual level understandable by the recipient. Tailored print communications (TPCs) may be an appropriate medium for cancer risk communication (CRC). TPCs are more refined than targeted communication materials. They are print materials created especially for an individual on the basis of knowledge about that person. The goal is to provide individually relevant and appropriate information. This review examines the nature of TPCs, assesses the use and potential of TPCs for the purpose of CRC, and highlights new directions in CRC. Articles dealing with TPCs were located by searching the MEDLINE® and PsychInfo® databases and seeking in-press articles. TPCs were identified for several areas of CRC, including dietary change, smoking cessation, mammography use, hormone replacement therapy, health risk appraisal, and genetic susceptibility to cancer. Although TPCs have been used in a number of different behavioral areas, they have not yet achieved their potential for CRC. The use of TPCs in the communication of cancer risk shows great promise, however, particularly as knowledge evolves regarding both the nature of risk and the most effective tailoring of health communication messages. [Monogr Natl Cancer Inst 1999;25:140–8]

Methods

We conducted MEDLINE® and PsychInfo® searches of the literature and identified fewer than 30 relevant articles on tailored print communications (TPCs). Here, we focus primarily on in-press or published articles that reported randomized studies of TPCs. We noted any mention of cancer risks as a tailoring or measurement variable in the study and examined study outcomes, especially related cancer risk variables. Because the focus of most of these articles was not on manipulation of cancer risks or on risk perceptions, the information that can be gleaned is often suggestive and is based on inferential rather than direct evidence, yet, in a larger sense, all these articles deal with cancer risk communication (CRC) because they focus primarily on reducing the chances of getting cancer or dying of it by changing cancer-related behaviors or engaging in cancer-protective behaviors.

TPCs were identified in several main areas of CRC: dietary change, smoking cessation, mammography use, hormone replacement therapy (HRT), and health risk appraisal (HRA). In addition, the amount of literature on genetic risk for cancer susceptibility is increasing and relevant as well because the communication is inherently individualized, personalized, and about risk.

Previous reviews by Strecher *(1)* and Skinner et al. *(2)* examined the first 12 studies published on TPCs. They concluded that, compared with their nontailored counterparts, TPCs have been consistently better remembered, read, and perceived as relevant, credible, or both. Evidence also exists that they are more effective for influencing health behaviors. Not all the studies showed main effects; in general, however, those that did not have main effects found important interactions. This article is not intended as a comprehensive review of TPCs, given the recent reviews. Rather, we were interested in identifying the more recent literature and examining the lessons learned and the research needed to advance the science of CRC. We sought evidence that the authors had tailored on risk or on risk perceptions or had discussed reactions to the TPCs in terms of altered risk perceptions, perceived susceptibility, or the like. Beeause of the great variability in populations, outcomes, and strategies and of our primary interest in the risk aspects of TPCs, we did not attempt a meta-analysis.

Background

The scientific calculation of risk is inherently statistical. For individuals, risk is personal. CRC has been defined as an interactive process of exchange that includes multiple messages about the nature of cancer and related risks *(3)*. To be effective, CRC must traverse a great conceptual distance—from probabilities, based on mathematics derived from populations, to the calculation of individual risk, and then to the accurate perception of personal risk by individuals. Individuals' perceptions about risk are influenced by many forces known only imperfectly by scientists. Ultimately, perception of risk should translate into knowledge or action. Stated more concretely, the chance that "a woman" will get cancer is far less salient than the likelihood that "I" will get cancer. This discrepancy may help to account for optimistic biases, the tendency for some people to discount their personal chanees of either getting a disease or experiencing something else that is undesirable *(4)*.

A real potential exists to improve health in the correct understanding of personal risk. It can increase knowledge, facilitate decision making, motivate new behaviors, and change existing behaviors *(5)*. Patients would benefit from health professionals who understand personal risk and have the time to communicate it to the patient on a one-on-one level.

A skilled, human risk advisor would take into account the host of individual factors that affect how people perceive risks and respond to risk communications, as well as age, cultural, and

Affiliation of authors: Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.

Correspondence to: Barbara K. Rimer, Dr.P.H., National Institutes of Health, Executive Plaza North, Rm. 242, MSC 7338, Bethesda, MD 20892-7338 (e-mail: brimer@nih.gov).

See "Notes" following "References."

risk-appropriate recommendations. In contrast, mass-produced materials on health education can provide information about general cancer risks, but they cannot take optimal advantage of the opportunity to communicate with individuals. They fail to recognize, as Maibach and Holtgrave *(6)* stressed, that the audience is not one undifferentiated group.

Whereas the personal relevance and accuracy of a risk estimate are important, so is the way it is presented. The manner in which cancer risk information is presented and how risk estimates are expressed may influence recipients' interpretations *(5)*. People need more than accurate, personal risk estimates *(7)*. They also must know how they, as individuals, can reduce their risks *(8)*.

CRCs should include, where appropriate, objective, absolute risks—for example, of getting cancer, dying of cancer, or having a mutation—and the objective risk reduction associated with recommendations for behavior changes. At present, only a small number of validated cancer risk models are available. Although these models yield apparently objective and quantitative risk estimates, tailoring carries both the opportunity and the obligation to help the individual reader understand the evidence underlying the model, as well as how a model-based estimate applies and does not apply to him or her. Any discussion must convey the confidence interval around the estimate. Such communications also should respond to people's subjective risk perceptions *(9)* and attempt to correct misperceptions. A person's subjective risk may be more important than his or her objective risk in determining better health behavior. On the one hand, a need exists for highly specific information about risk, presented in a way that takes into account the relevant characteristics that influence the understanding and use of the information. On the other hand, traditional booklets and brochures are limited in their capacity to individualize risk communication.

To resolve this dilemma, health communicators have begun to use TPCs. TPCs can accomplish this multilevel communication because individualized strategies are appropriate when the information must be individualized to be meaningful. Thus, cancer risk information can include other, individually appropriate content as well as specific quantitative estimates. Probability estimates can be accompanied by both an explanation of how the risk estimate was derived and what it means. It can include information about how to reduce the risk *(10,11)*. CRC also must reflect the fact that some people do not want to receive quantitative probability estimates. Ideally, risk communication involves truly interactive communication and shared decision making about recommended risk management strategies *(6)*.

TPCs are print materials that are created especially for an individual. They are based on information about that person *(2)* and provide individually relevant and appropriate information *(12)*. The information can come from many sources (e.g., self-administered or interviewer-administered questionnaires, medical records, and encounter data) *(13)*. The diffusion of TPCs has been facilitated by advances in computer technology and a growing number of software programs that can be used to create TPCs. In addition, the genetics revolution is providing us with the tools to provide better estimates of genetic risks.

Everywhere we look, the concept of mass customization is being applied *(14)*. Increasingly, the consumer expects to be treated as an individual. Mass marketing has become more and more individualized, and the evidence suggests consumer acceptance of tailored products. TPCs are part of a global information revolution. Robinson et al. *(15)* recently highlighted the potential advantages of the "new media" for improving health, in part, by tailoring information to the needs of users. Tailored communications may be particularly important in helping consumers make personal decisions about health care choices. This ability is especially important in an era when the profusion of choices and the complexity of treatments may leave consumers at a disadvantage *(16)*.

TPCs are suited ideally to communicating cancer risks. We provide an overview of TPCs, with a focus on their potential for CRC. We briefly will review the studies conducted to date, with a focus on their utility for CRC. We also will highlight new directions in CRC and the potential of TPCs for use in CRC. Another review in this monograph focuses on interactive health communications *(17)*.

How Tailoring Is Done

Tailored materials are more refined than targeted materials. The latter may segment the population into high-risk groups—for example, on the basis of race, a health condition, or age—and then create materials especially for the group. Evidence suggests that targeted materials are effective *(17)*. But the messages are not individualized as they are for tailored materials. Researchers have hypothesized that greater effect could be achieved through tailoring.

Tailoring should be informed by theory and what is known about the determinants of the behavior *(18)*. For example, to create a TPC about smoking cessation, the creator would need at least the following data: patient's name, sex, current smoking status and intentions, responses to items about the pros and cons of change, and responses to the processes of change items *(19)*. Ideally, level of addiction, confidence in ability to quit, and related variables also would be assessed. The selection of tailoring variables should be grounded in theory and understanding of the behavior. Some tailoring algorithms have used variables such as risk perceptions, motivations to change, quantitative risk estimates, and self-efficacy. This tailoring is especially appropriate when models such as Precaution Adoption are used.

To accurately assess risk and risk perceptions in TPCs, researchers should use qualitative and quantitative measures of risk as recommended by Weinstein et al. *(20)*. These measures should include how likely one is to get breast cancer, for example. One would want to ask this question, using both qualitative descriptors, such as "How likely are you to get *x*?" as well as "On a scale from 1 to 100, what is the chance you will get *x*?" In addition, both objective and subjective risks should be assessed. The assessment should be specific to a particular risk or hazard, and the appropriate reference group should be used.

Tailoring of print materials may be done quite simply, using a few variables such as race and previous health behavior. Or it may be done by using complex combinations of variables that result in millions or billions of possible combinations.

TPCs come in many forms. The first reported TPCs were variants on tailored, personalized letters and feedback reports *(21–23)*. More recent studies (most in press or ongoing) are using a wider array of approaches, including tailored calendars *(24)*, tailored birthday cards and newsletters *(25)*, tailored brochures *(13)*, tailored tip sheets in combination with feedback reports *(26)*, and combinations of personalized feedback reports and stage-matched manuals *(19)*. Investigators have used different approaches when programming the tailored communications

algorithms, including combinations of spreadsheets, databases, Microsoft Word, and Pascal. Unfortunately, at present, no "off the shelf" packages are available for producing tailored communications.

Briefly, print message tailoring is accomplished by programming a computer application, such as a word processor, to select and combine message elements into a printed message. The message elements are part of a message library that contains all possible responses for different conditions. The basis for the selection is the variables and their values in an individual's data record, using information drawn from such sources as interviews, medical or pharmacy records, and observations. Message elements are usually sentences and paragraphs, as well as images such as drawings and graphs. Empirically based decision rules for selecting the appropriate message from a library of options are a feature of the expert-like systems that are used to produce TPCs. Tailored risk communication might be seen as simply a subset of TPCs, i.e., those communications dealing with risk, but this statement would oversimplify the range and complexity of tailoring opportunities associated with risk.

How Tailoring Works

It is not known exactly how tailoring exerts its effect. Data show that tailored materials are more likely to be read, retained, remembered, and rated more highly in comparison with nontailored materials. Under Petty and Cacioppo's *(27)* Elaboration Likelihood Model, the increased attention and retention might lead to better processing of the information and, thereby, to better outcomes. Such outcomes are consistent with Rothman and Schwartz's *(28)* observations about the importance of the accessibility of risk information. Features of both the person and the environment affect accessibility. The effect of TPCs also is consistent with the Transtheoretical Model *(29)* as the guiding theory; many tailored materials address issues pertinent to a particular stage of change without extraneous information. They are perceived as more relevant and are more attended to by the recipient and, ultimately, facilitate more behavior change. In the case of CRC, following the Precaution Adoption Model of Weinstein et al. *(20)*, the effect may be achieved because the TPCs are directed at a person's position along a behavioral continuum that reflects response to a hazard and overcomes barriers. Or messages might be designed to increase perceived susceptibility consistent with the Health Belief Model *(30)*.

Because the reported studies have used several different theories, variables, and formats and generally have not used factorial designs, it is not possible to answer some of the most interesting and important questions about how tailored materials exert their effect, much less the role of risk communication in producing the effect. Moreover, most TPC research reported to date has relied on theories, such as the Stages of Change Model, in which risk is not a major construct. Several of the TPC studies currently in the field are using theories, such as the Precaution Adoption Model, in which risk is central. Thus, much more information should be available in the future about the utility of TPCs for CRC.

Overview of Research Results

In the next sections, we review research results from the TPC literature. Because two recent reviews *(1,2)* have focused on many of the articles reviewed here, our summary is brief. In Tables 1 and 2, we highlight the study outcomes and the role of risk in those studies.

Table 1. Evidence for the effectiveness of tailored print communications (TPCs)

Characteristic	Significant impact* Yes	No
TPCs are more likely to be read, recalled, and rated more highly.	Brinberg and Axelson *(31)* Campbell et al. *(32)* Brug et al. *(22)* Lipkus et al. *(33)* Rimer et al. *(25)*	
TPCs have a significant effect on decreasing dietary fat intake.	Bowen et al. *(34)* Brinberg and Axelson *(31)* Campbell et al. *(32)* Brug et al. *(22)* Brug et al. *(23)*	
TPCs have a significant effect on fruit and vegetable intake.	Brug et al. *(23)*	Campbell et al. *(32)* Brug et al. *(22)*
TPCs have a significant main effect or subgroup effect on smoking cessation.	Strecher et al. *(35)* Lipkus et al. *(33)* Dijkstra et al. *(36)*	Curry et al. *(37)* Dijkstra et al. *(38)*
TPCs have a significant main effect or subgroup effect on use of mammography.	Skinner et al. *(21)* Rakowski et al. *(26)*	Rimer et al. *(25)* Drossaert et al. *(39)* Meldrum et al. *(40)*
TPCs improve woman's decisions about hormone replacement therapy.	O'Connor et al. *(41)*	
TPCs can improve behavior change in response to a health risk appraisal.	Kreuter and Strecher *(42)*	

*Investigators and reference numbers are listed in columns.

Dietary Change

To date, more TPC studies have been conducted in the area of dietary change than in any other area. However, assessment of risk perceptions and tailoring on the basis of risk played only a minor role in these studies. Campbell et al. *(32)* measured perceived susceptibility to diet-related diseases but did not report findings. One of the assumptions of Brug et al. *(22)* was that TPCs would heighten recipients' awareness of cancer risks and lead to other changes that would result in dietary behavior changes. Thus, feedback about dietary practices was designed accordingly. Nevertheless, the results have been impressive—significant effects on decreasing dietary fat intake in studies by Brinberg and Axelson *(31)*, Bowen et al. *(34)*, Campbell et al. *(32)*, and Brug et al. *(22,23)*. These outcomes were achieved with a variety of TPCs, including brochures *(31)*, personalized feedback *(34)*, and a tailored nutrition packet *(32)*. The results for increasing fruit and vegetable consumption with the same interventions were mixed—one study *(23)* found significant effects, while two other studies *(22,32)* did not.

As a group, these studies have demonstrated that tailored letters can change dietary behavior, although the results were more consistent for dietary fat than for fruit and vegetable consumption (*see* Tables 1 and 2). The authors also found that TPCs, compared with generic materials, were more likely to be read, were rated higher, and resulted in greater information re-

Table 2. Use of risk messages and/or assessment of risk outcomes in TPCs*

Authors (reference No.)	Risk message and/or impact
Campbell et al. *(32)*	Beliefs about perceived susceptibility to diet-related diseases were measured but not reported.
Bowen et al. *(34)* Brug et al. *(22)* Brug et al. *(23)*	TPCs heightened awareness of cancer risks as related to diet, but no risk-related outcomes were presented.
Curry et al. *(37)*	Health concern was a tailoring variable, but no outcome data were presented.
Strecher et al. *(35)*	Risk for smoking-related diseases was a tailoring variable.
Dijkstra et al. *(38,43)*	Smoking cessation was tailored on outcome and efficacy expectancies; perceived health consequences were not affected.
Lipkus et al. *(33)*	As a facilitator for quitting smoking, improved health was a tailoring variable but was not related to quitting.
Skinner et al. *(21)*	Perceived susceptibility was not affected by the intervention.
O'Connor et al. *(41)*	Women who received tailored decision aids were more likely to be aware of the risks and benefits of HRT†.
Kreuter and Strecher *(42)*	Enhanced HRA‡ feedback included risk information and individually tailored behavior change information. Patients who received enhanced HRA feedback were 18% more likely to change at least one risk behavior. Patients who overestimated their cancer risks became more accurate.

*TPCs = tailored print communications.
†HRT = hormone replacement therapy.
‡HRA = health risk appraisal.

call. One important similarity underlying these studies was an assumption that awareness of one's cancer risk behavior is necessary for change and that personalized, individualized information about one's own behavior can provide motivation for change. In large measure, the studies were successful. However, the published reports do not permit readers to draw conclusions about the role of risk or risk perceptions in producing the outcomes. Nor, as in the other topic areas, can the reader assess to what extent the overall quality of the TPC or particular elements, such as information, format, or style, affected outcome.

Smoking Cessation

The amount of literature on the use of TPCs to increase smoking cessation is increasing. The smoking studies generally have used the Stages of Change Model as the theoretical foundation, but they have varied greatly in how tailoring was done as well as in style, format, number of tailoring variables, and other elements. As in the area of dietary change, it is difficult for the reviewer to judge the quality of the TPC. Tables 1 and 2 show that the effect of TPCs on smoking cessation has been mixed, with positive findings in three studies *(33,35,36)* and negative findings in two studies *(37,38)*.

Risk communication could be a strong component of smoking cessation interventions, but not all studies were tailored on risk perception variables or even measured such variables. Studies by Strecher et al. and Dijkstra et al. are exceptions. Strecher et al. *(35)* used risk for smoking-related diseases as a tailoring variable, and messages included the perceived threat of smoking-related diseases. No data were reported with regard to changes in perceived risk. Dijkstra et al. *(36,38,44)* assessed different questions related to the effect of TPCs on smoking cessation. Of particular relevance to risk communication is a study that focused on using computer-tailored feedback to change cognitive determinants of smoking among smokers who said they were planning to quit in 6 months. The authors compared computer-generated letters on the outcomes of smoking cessation with computer-generated letters, using efficacy information versus control conditions *(38)*. They focused on social learning variables, such as self-efficacy and outcome expectancies. Although some variables, such as perceptions about environmental and social consequences, were changed as a result of the computer letters, perceived health consequences—a dimension of risk perception—were not affected.

Lipkus et al. *(33)* assessed the effect of iteratively tailored birthday cards and newsletters among low-income, primarily African-American smokers. They tailored the cards and newsletters on the basis of variables from the Stages of Change Model as well as barriers to quitting and other variables. Three intervention groups were identified: provider prompting only, tailored print, and tailored print and telephone counseling. A highly significant quitting effect was seen for the TPC group: 33% of those who received TPCs quit smoking compared with 13% in the prompting only group and with 19% who received telephone counseling as well as the TPC. The effect was especially strong for contemplators and black men. Although improved health was a motivator for quitting and therefore a tailoring variable, risk perceptions were a minor part of both the intervention and evaluation and did not seem to affect quitting behavior.

Thus, several TPC studies have focused on smoking cessation. Several different kinds of TPCs were used in these studies, including personalized feedback *(37,43)*, a tailored letter *(35,38,44)*, and tailored birthday cards and newsletters *(33)*. Some investigators used pro or con measures and measures of risk perception and then tailored on these variables. Although risk perceptions, health concerns, or both were tailoring variables in some studies, they did not appear to be a major focus, and it is not possible to determine how risk tailoring affected the outcomes. Two of the studies had main effects on longer term abstinence. An additional study showed significant interactions (*see* Tables 1 and 2). However, the variations in smoker characteristics make comparisons difficult.

Mammography Use

In several studies, TPCs were used as part of a strategy to increase mammography use. In most of these studies, risk was not a major variable, although, in some cases, the TPC was designed to reduce women's unrealistic perceptions of cancer risk. Skinner et al. *(21)* conducted one of the first studies in this area and found no main effects on mammography use but significant interactions, with special benefits for black women and those with incomes less than $26 000. The tailored letter did not result in differences in perceived susceptibility, Skinner's risk perception variable *(21)*.

Two European studies *(39,40)* used tailored leaflets and invitation letters to increase participation in mammography. Neither study found an improvement in participation among women who received the tailored letters and leaflets. But mammography rates were uniformly high and these were very minimal TPCs.

In the United States, Rakowski et al. *(26)* combined stage

matching and tailoring to create tip sheets and tailored letters to increase mammography use among 1864 women in a health maintenance organization (HMO). Stage-matched, tailored materials significantly outperformed a no-materials group and a standard intervention group: 63.6% of the women in the stage-matched group received mammograms compared with 54.9% in the no-materials group and with 58.5% in the standard intervention group. Tailoring was not based on risk perceptions or communication of actual risks.

In a study in which TPCs were used in a low-income, primarily African-American population, Rimer et al. *(25)* found an effect of tailored birthday cards and tailored newsletters for Pap tests and overall cancer screening but not for mammography. However, mammography use was already well above the national average in all study groups, so there was little room for improvement.

The mammography-related TPCs are promising in spite of inconsistent results. This area might benefit from more inclusion of and tailoring on risk variables, as there are quantitative models of breast cancer risk. Some current studies are doing just that. For example, in an ongoing study using TPCs to improve women's decision making about mammography, Rimer and colleagues *(13)* are giving women their Gail Model scores with additional information to shift their risk perceptions, if appropriate. Balance sheets are used to help women examine the pros and cons of mammography. Women are informed of the factors that contributed to their Gail Model scores. This approach is consistent with Rothman and Kiviniemi's *(7)* recommendation that risk communicators provide information on probabilities and also on causes and consequences.

Hormone Replacement Therapy

Few areas of medicine exist in which risk communication is more challenging than HRT. Providers must communicate, and women must understand the trade-offs of risks and benefits for multiple diseases as they consider a lifelong HRT regimen. O'Connor et al. *(41)* applied the exciting technology of decision aids in helping women to make personal decisions about HRT use. In a randomized clinical trial, the authors compared the efficacy of a general educational pamphlet with a tailored decision aid. Women who received the decision aids showed significant improvements in a number of areas, including realistic personal expectations of the risks and benefits of HRT. Women who received the decision aids were more likely to know HRT benefits and risks, to be aware of the importance of the risks, and to say they made an informed choice (*see* Tables 1 and 2).

Health Risk Appraisal

Kreuter and Strecher *(42)* conducted a randomized trial to compare the effects of typical HRA feedback with enhanced HRA feedback. The enhanced feedback was individually tailored on the basis of the Health Belief Model and Stages of Change Model. The feedback included two components: risk information and individually tailored behavior change information. Overall, patients who received enhanced HRA feedback were 18% more likely to change at least one risk behavior than the patients who received typical HRA feedback. The enhanced feedback promoted changes in some behaviors but not in others, as determined by a baseline assessment and a 6-month follow-up. Moreover, the enhanced HRA was effective in changing perceived risks. Patients who had underestimated their stroke risks and those who had overestimated their cancer risks both became more accurate *(45)*. This is one of the few studies to demonstrate a causal link between risk perceptions and behavior. Smokers who had accurate perceptions of stroke risk at baseline were significantly more likely to have quit smoking at follow-up than were those with optimistic biases. These results suggest that tailored communications may be effective for some behaviors when based on risk information and behavior change information.

Genetic Susceptibility to Cancer

One of the areas of CRC that provides the greatest challenges is cancer genetics. Genetic risk is inherently probabilistic, and, as Wroe et al. *(46)* showed, perceived risk may be a more important variable for genetic testing than for cancer screening and other health screening. Lloyd et al. *(47)* pointed out that the effectiveness of genetic counseling depends on attenders' comprehension of risk communication. They also suggested that attention should be given to factors that facilitate comprehension and recall of risk estimations. In the literature related to genetic susceptibility for cancer, a number of authors have assessed people's perceptions about their cancer risks, in addition to other measures *(48)*. Most of these authors have found that people overestimate the chances of having a mutation and of getting cancer. Moreover, when faced with the possibility of being tested for genetic susceptibility, most people likewise focus on the benefits rather than on the risks or limitations. Thus, this area might benefit from the use of individualized educational strategies such as TPCs.

Few investigators have attempted to experimentally control the risk communication or even to assess people's reactions to risk estimates. We will focus here on this aspect of genetic risk studies because they share a common core with TPC studies—the communication of individualized risk estimates. Although these studies are not strictly TPC studies, they offer lessons to those interested in considering the appropriateness of TPCs for communieating cancer risks in the area of genetics.

In one of the few experimental studies, Lerman et al. *(49)* conducted a randomized trial in which high-risk relatives of patients with breast cancer were assigned to either a general health education intervention or one in which they received individualized Gail Model scores and related information. Women who received risk counseling and printed risk estimates were significantly more likely to improve their risk comprehension. However, women in both groups continued to overestimate their lifetime risk of breast cancer. Women with high levels of baseline anxiety were least likely to benefit from the intervention. Lloyd et al. *(47)* and Watson et al. *(50)* studied women who attended genetics clinics versus general practice clinics to assess the effect of personalized risk estimates. They found that most women overestimated their breast cancer risk; 66% could not accurately recall their own risk even after being provided printed estimates.

The difficulty in recalling risk estimates may result from how information is presented. Hallowell et al. *(51)* studied preferences for risk information and found that 73% of women wanted quantitative risk estimates; however, in over 40% of cases, the risk information was not presented in the counselee's desired format. In generic health education materials, it is not possible to respond to format preferences. TPCs offer the potential to tailor on both the risk estimate and the manner in which it is presented.

In a study being conducted at Duke University Medical Center through the Breast Specialized Program of Research Excellence (SPORE), women were asked whether they wanted a qualitative risk estimate, a quantitative risk estimate, or both. Baseline data indicate that the majority of women wanted both estimates *(8)*. However, some women want only qualitative estimates, and this estimate is what they are given until their counseling session. A real mismatch could occur if women wanted qualitative estimates and were provided with numbers. The ability to respond to the needs, readiness, and preferences of individuals is an inherent benefit of TPCs. Fig. 1 shows an example of a woman who received a quantitative estimate of risk.

Together, these studies show that most people want quantitative risk estimates, but these estimates are not enough to encourage accurate recall. Moreover, the uncertainties associated with the results of genetic testing are especially challenging. Welch and Burke *(52)* gave the example of a physician who needed to communicate the concept: "You have a 75% risk of developing colon cancer." TPCs may be a good vehicle for communicating genetic risks because they can provide individualized estimates in a format that is acceptable to individuals.

Discussion

We have reviewed reports of 21 published or in-press studies in which the first TPCs were evaluated in randomized trials. As we discussed, TPCs have been used in a number of different behavioral areas, including primary and secondary cancer prevention as well as in cancer genetics, HRA, and decision making. Because the studies were so different from one another, with very different outcomes, we did not attempt a meta-analysis. Nevertheless, we can offer several conclusions.

First, as shown in Table 2, with reference to the theme of this review, TPCs have not achieved their potential in CRC. Whereas all the topics covered by TPCs can be considered CRC, at least broadly, little investigation has been done of how risk mediates the outcomes achieved. Even when authors said they were using a model that incorporated risk perception or objective risk measures, they rarely presented results specific to these variables. Two studies provide reason to be optimistic about the potential role of TPCs in CRC. Kreuter and Strecher *(42)* demonstrated that patients who received enhanced HRA feedback in comparison with standard feedback became more accurate in their estimates of cancer risks. O'Connor et al. *(41)* showed that decision aids about HRT resulted in patients who were more aware of the risks and benefits of HRT.

Clearly, TPCs are rated more highly than their generic counterparts. This finding is important, and other investigators should use process measures as part of their evaluations. All the TPC studies conducted on dietary fat have found significant effects, whereas the results for fruit and vegetable consumption have been equivocal. However, in several of the studies, the investigators were attempting to affect multiple behaviors. Thus, it is not clear whether the lack of effect is a result of behavioral overload or if fruit and vegetable consumption is a more difficult behavior to influence through TPCs. The two smoking cessation studies that used well-developed TPCs without auxiliary print materials and did not self-select particularly difficult populations *(33,35)* found significant main effects or interactions. This finding is especially important given the difficulty of influencing smoking behavior. The results on mammography also are equivocal, but the studies may have suffered from ceiling effects. Finally, in the areas of HRT and HRA, the results were positive.

In the studies reported in this review, it was not possible to determine whether, and to what extent, quality differences in the construction of the TPCs might have affected study outcome. While a rose is a rose is a rose, to paraphrase Gertrude Stein, the same cannot be said for TPCs. It is possible that negative results are a result of inadequacies in the TPC itself.

The TPCs reviewed here varied greatly in design, format, length, and execution. Some may have been executed better than others. No reason exists to believe that a letter with a few tailored elements would be as effective as a brochure with information and graphics tailored to the reader.

Thus, on balance, more positive than null results were found. This finding is true especially if one focuses on the more recent studies. The studies that are now in the field will be poised to answer many of the important questions about tailoring. Overall, the results of the studies reviewed here give us reason for both optimism and caution. The results are generally balanced; certainly no evidence of harm is found. However, more work is needed to determine how best to tailor and assess what problems and populations are most appropriate for tailoring interventions.

To advance the science of both tailoring and CRC, scientists should agree on the most pressing questions, develop a core set of common terms and measures, and put their codes, like genetic sequences, into a researcher-accessible website. TPC researchers can do a better job of describing what they do and of making their tailoring algorithms and formats available for examination. Science is built through iterative discoveries as well as giant leaps. Major advances in this field are likely to come from a concerted effort by scientists to collaborate to raise and answer the important questions.

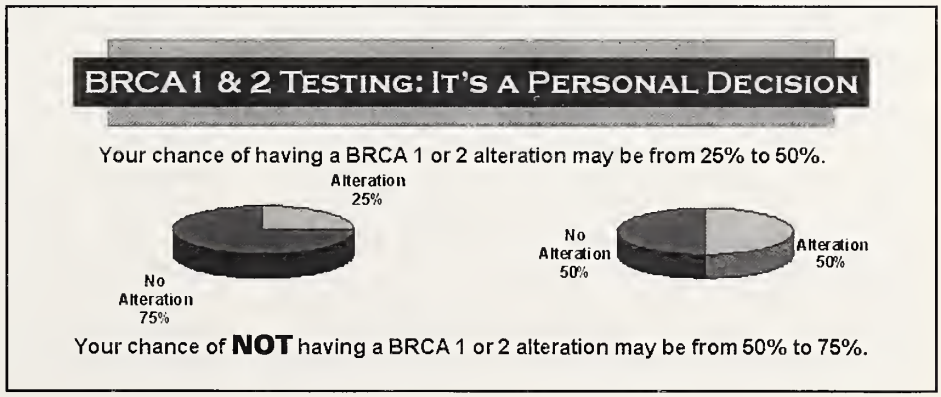


Fig. 1. Communicating personal risk information.

In the next sections, we summarize some of the lessons learned about TPCs for health practitioners and suggest several areas of research that should be pursued to answer important questions about TPCs.

What Can Practitioners Do Today With Tailored Risk Communication?

Several reviews have highlighted the value of personalized letters and reminders *(53)*. The previous discussion shows, on balance, a favorable effect of TPCs. The collective experience generally has been with tailored letters, booklets, newsletters, and even birthday cards. To date, it is not clear how much tailoring is needed. In many cases, less may be more. Even a Post-it note with an individualized message can increase the salience of health messages. Wherever possible, physicians' offices, health departments, and other organizations should use tailored, personalized letters rather than, for example, "Dear Patient" letters. Any word processor is capable of producing a personalized letter of advice or a cover letter for a standard brochure that is addressed to the patient by name. A more intensive level of tailoring might be to use the data about a patient to insert a paragraph or sentence in the cover letter that points to a particular section of the brochure ("Be sure to read the section beginning on page *x* of the enclosed booklet").

Another approach would be a risk-specific insert for the front of a generic booklet. The practitioner would create inserts for, say, three levels of risk and insert one by hand before giving or mailing the booklet to the individual. The insert could include areas to hand write details such as laboratory values and risk categories. New products, such as National Cancer Institute's "Risk Disk," provide ways to calculate an individual's probability of being diagnosed with breast cancer. More such products will be available in the future as risk algorithms become more commonplace. At the very least, tailored letters will be needed to highlight or reinforce such information.

Strategy Choices

Practitioners should look to theory and to evidence-based scientific literature to guide strategy choices based on evidence. The literature now testifies to the benefits of targeted and tailored materials over generic ones. At this stage, parsimonious approaches to tailoring, such as letters, cards, and newsletters, are supported by the literature *(53)* in a number of areas. Some evidence shows that multiple tailored communications are more effective than single communications.

Practitioners should apply what is already known about risk communication in developing printed materials. Paring down the information to what people need should be a fundamental step in CRC. Fischhoff *(54)* has stressed the importance of understanding what is really relevant to a behavior or decision. Finally, wherever possible, researchers should evaluate what they do in tailoring. By doing so, they will improve their own work and advance both the state of the science and the state of the art.

A Research Agenda for Tailored Print Communications

The TPCs currently being tested are more advanced in terms of theory, format, and tailoring strategy than some of those reported here. Moreover, researchers are applying the strategy of tailored communication to areas about which there has been controversy, such as mammography for women in their 40s, colorectal cancer screening, HRT, and tamoxifen use for breast cancer risk reduction. In such cases, TPCs might serve an important role in the decision-making process not only by communicating critical information about risks and benefits but also by helping users to weigh their own perceptions of risks and benefits. In addition, researchers, such as those participating in the Cancer Information Service Research Consortium (CISRC), are using appropriate designs to answer important questions (e.g., whether iterative tailoring is both effective and cost-effective). A number of important research questions about CRC could be examined using TPCs.

1) To advance CRC, TPC researchers should be encouraged to provide the following information consistently: theory used to inform the intervention and how the TPC is expected to exert its effect, as well as summary of risk measures, risk messages, and changes in risk variables as a result of exposure to TPCs. In addition, researehers should follow the recommendations put forward by Robinson et al. *(15)* for reporting studies of interactive communications. They provided detailed recommendations for process and outcome evaluation measures.

2) Researchers should be encouraged to assess people's reactions to tailored risk communications and provide data about changes in variables such as perceived risk, susceptibility, and worry. The role of risk as an intervening variable affecting responses to TPCs should be investigated and then reported. Most current reports do not provide enough information to assess how changes in risk perceptions may have affected study outcomes.

3) As Weinstein and Sandman *(55)* advised, measures of risk comprehension should be obtained, and careful attention should be given to the criteria by which messages are to be evaluated. This process should be a part of the reporting for TPCs when risk is a major component.

4) TPCs should be tested in factorial designs or in smaller scale and more controlled laboratory studies to examine how to communicate risk estimates and other risk-related information. This testing is especially important as quantitative estimates of risk are increasingly available. Moreover, such designs could help to elucidate basic mechanisms. As de Vries and Brug *(56)* pointed out, we still do not know why tailored health education is more effective than general health education because no study was designed to answer this question. In this area, as in many other areas of CRC, we need to understand basic mechanisms.

5) TPCs can be used to improve the communication of cancer risk by refining messages so that they are appropriate for particular exposures and audiences. Such tailoring also could use increasingly available data on neighborhood exposures and behaviors.

6) More research is needed to assess the extent to which the efficacy of TPCs may be enhanced by combining them with other strategies such as companion materials for health providers or other adjuncts such as telephone counseling. The results to date have been inconsistent. For example, Rimer et al. *(25)* found a benefit of adding telephone counseling to TPCs for cancer screening. But Lipkus et al. *(33)* showed that TPCs outperformed the counseling condition for smoking cessation.

7) Future research should consider how factors such as cultural context, emotions, personality, and utilities can be incorporated into TPCs to enhance their relevance *(46)*. Recently, the Institute of Medicine *(57)* made a series of recommendations about CRC in following up on an analysis of thyroid cancers related to iodine-131 exposure. Among other topics, the report called for research to determine how to communicate risk infor-

mation so it is perceived as personally relevant. Although the data suggest that TPCs are more likely than generic materials to be perceived as personally relevant, more research is needed to determine how to enhance perceived relevance. More understanding of people's mental models may be helpful in this area *(54)*.

8) More research is needed to refine the creation of tailored decision aids, especially those that are appropriate for people with a high school education or less as well as ethnic minorities.

9) A real need exists for statisticians to collaborate with behavioral scientists to develop more refined techniques to evaluate the effect of TPCs on the basis of every participant getting a unique intervention.

The Future

Chamberlain *(58)* argued that, instead of addressing mass audiences, tomorrow's communicators need to learn how to send multiple messages through myriad channels, tailored in such a way that the content appears to be (or can evolve into) an individual message for the receiver. Risk communication research should be undertaken with the fundamental assumption that the resulting message is intimately tied to the individual—to age, interests, education, experience, health needs, and perceptual style. Such tailoring should accommodate preference for graphical information. This accommodation can be done with greater detail than a health professional talking one on one with a patient.

Is there benefit in telling a smoking security guard that for every one of his or her colleagues who will draw a gun in self-defense, one hundred will die of a smoking-related illness before reaching retirement?

Can an accomplished golfer be persuaded by comparing the chances of making a hole in one on a familiar golf course with the chances of getting skin cancer from playing without a hat?

Even at a less deeply detailed level, research questions and directions will be enriched by close interplay between understanding individuals' perceptions of risk and pushing back the limits on how we can individually operationalize our understanding.

Some people seem to conceptualize space better; others seem to perceive best in terms of time. Weinstein et al. *(20)* suggested that describing different risks in terms of time intervals might be an effective way of communicating risk probabilities. This approach may be especially important when communicating population risks. Some people use rules of thumb and rough guidelines in decision making, while others go by the book. Some prefer statistical arguments; others prefer narrative and synecdoche. With tailoring, any combination of those attributes and others could be the basis for a communication. Tailoring research must become sufficiently multidimensional so that it can illuminate some of the more obscure links among personality, cognition, message characteristics, and action.

How to construct a parsimonious risk-perception profile that can drive state-of-the-art tailored communication methods is a question that needs to be asked, researched, and asked again. Both content and context will evolve along with our understanding of the biology of risk, with the psychobiology of risk perception, and with the technology of tailoring.

Researchers should examine the interrelationships between objective and subjective risks. For example, many women overestimated their subjective risk of breast cancer in comparison with their objective risk based on Gail Model scores *(59)*. However, smokers may underestimate their risks for lung cancer. Providing tailored feedback on these mismatches could be a potential value of TPCs.

Like other new technologies, which are becoming increasingly available as a result of the communications revolution, TPCs are not a panacea. But the thoughtful use of theory-driven TPCs, built on a foundation of evidence from diverse fields (including the communication and behavioral sciences), could improve the art and science of CRC *(60)*.

References

(1) Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. Patient Educ Couns 1999;36:107–17.

(2) Skinner CS, Campbell MK, Rimer BK, Curry S, Prochaska JO. How effective is tailored print communication? Ann Behav Med. In press 1999.

(3) National Research Council, Committee on Risk Perception and Communication. Improving risk communication. Washington (DC): National Academy Press; 1989.

(4) Weinstein ND. Optimistic biases about personal risks. Science 1989;246:1232–3.

(5) Bottorff JL, Ratner PA, Johnson JL, Lovato CY, Joab SA. Communicating cancer risk information: the challenges of uncertainty. Patient Educ Couns 1998;33:67–81.

(6) Maibach E, Holtgrave DR. Advances in public health communication. Ann Rev Public Health 1995;16:219–38.

(7) Rothman A, Kiviniemi MT. Treating people with health information: an analysis and review of approaches to communicating health risk information. Monogr Natl Cancer Inst 1999;25:44–51.

(8) Rimer BK, Sugarman J, Winer E, Bluman LG, Lerman C. Informed consent for BRCA1 and BRCA2 testing. Breast Dis 1998;10:99–114.

(9) Palmer CGS, Sainfort F. Toward a new conceptualization and operationalization of risk perception within the genetic counseling domain. J Genet Couns 1993;2:275–94.

(10) National Cancer Institute, Office of Cancer Communications. How the public perceives, processes and interprets risk information: findings from focus group researeh with the general public. Bethesda (MD): National Cancer Institute; 1998 June Report No.: POS-T086.

(11) Lipkus IM, Hollands JG. Using graphics to communicate disease risk. Monogr Natl Cancer Inst 1999;25:149–63.

(12) Velicer WF, Prochaska JO, Bellis JM, DiClemente CC, Rossi JS, Fava JL, et al. An expert system intervention for smoking cessation. Addict Behav 1993;18:269–90.

(13) Rimer BK, Glassman B. Tailoring communications for primary care settings. Methods Inf Med 1998;37:171–7.

(14) Pine BJ. Mass customization: the new frontier in business competition. Boston (MA): Harvard Business School Press; 1993.

(15) Robinson TN, Patrick K, Eng TR, Gustafson D. An evidence-based approach to interactive health communication: a challenge to medicine in the information age. Science Panel on Interactive Communication and Health. JAMA 1998;280:1264–9.

(16) Gustafson DH, Robinson TN, Ansley D, Adler L. Consumers and evaluation of interactive health communication applications. Am J Prev Med 1999;16:23–29.

(17) Rimer BK, Orleans CT, Fleisher L, Cristinzio S, Resch N, Telepchak J, et al. Does tailoring matter? The impact of a tailored guide on ratings and short-term smoking-related outcomes for older smokers. Health Educ Res 1994;9:69–84.

(18) Rakowski W. The potential variances of tailoring health behavior interventions. Ann Behav Med. In press 1999.

(19) Velicer WF, Prochaska JO. An expert system intervention for smoking cessation. Patient Educ Couns 1999;36:119–29.

(20) Weinstein ND, Kolb K, Goldstein BD. Using time intervals between expected events to communicate risk magnitudes. Risk Anal 1996;16:305–8.

(21) Skinner CS, Strecher VJ, Hospers H. Physicians' recommendations for mammography: do tailored messages make a difference? Am J Public Health 1994;84:43–9.

(22) Brug J, Steenhuis I, van Assema P, de Vries H. The impact of a computer-tailored nutrition intervention. Prev Med 1996;25:236–42.

(23) Brug J, Glanz K, Van Assema P, Kok G, van Breukelen GJ. The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake. Health Educ Behav 1998;25:517–31.
(24) Kreuter MW, Vehige E, McGuire AG. Using computer-tailored calendars to promote childhood immunization. Public Health Rep 1996;111:176–8.
(25) Rimer BK, Conaway M, Lyna P, Glassman B, Yarnall KSH, Lipkus I, et al. The impact of tailored interventions on a community health center. Patient Educ Couns. In press 1999.
(26) Rakowski W, Ehrich B, Goldstein MG, Rimer BK, Pearlman DN, Clark MA, et al. Increasing mammography among women aged 40–74 by use of a stage-matched, tailored intervention. Prev Med 1998;27:748–56.
(27) Petty RE, Cacioppo JT. Attitudes and persuasion: classic and contemporary approaches. Boulder (CO): Westview Press, Inc.; 1996.
(28) Rothman AJ, Schwartz N. Construction perceptions of vulnerability: personal relevance and the use of experiential information in health judgments. Personality Soc Psychol Bull 1998;24:1053–64.
(29) Prochaska JO, DiClemente CC, Velicer WF, Rossi JS. Standardized, individualized, interactive, and personalized self-help programs for smoking cessation. Health Psychol 1993;12:399–405.
(30) Champion V, Foster JL, Menon U. Tailoring interventions for health behavior change in breast cancer screening. Cancer Pract 1997;5:283–8.
(31) Brinberg D, Axelson MA. Increasing the consumption of dietary fiber: a decision theory analysis. Health Educ Res 1990;5:409–20.
(32) Campbell MC, DeVellis BM, Strecher VJ, Ammerman AS, DeVellis RF, Sandler RS. Improving dietary behavior: the effectiveness of tailored messages in primary care settings. Am J Public Health 1994;84:783–7.
(33) Lipkus IM, Lyna PR, Rimer BK. Using tailored interventions to enhance smoking cessation among African-Americans at the community health center. Nicotine Tob Res. In press 1999;1:77-85.
(34) Bowen DJ, Tomoyasu N, Anderson M, Carney M, Kristal A. Effects of expectancies and personalized feedback on fat consumption, taste, and preference. J Appl Soc Psychol 1992;22:1061–79.
(35) Strecher VJ, Kreuter M, Den Boer DJ, Kobrin S, Hospers HJ, Skinner CS. The effects of computer-tailored smoking cessation messages in family practice settings. J Fam Pract 1994;39:262–70.
(36) Dijkstra A, De Vries H, Roijackers J, van Breukelen G. Tailored interventions to communicate stage-matched information to smokers in different motivational stages. J Consult Clin Psychol 1998;66:549–57.
(37) Curry SJ, McBride C, Grothaus LC, Louie D, Wagner EH. A randomized trial of self-help materials, personalized feedback, and telephone counseling with nonvolunteer smokers. J Consult Clin Psychol 1995;63:1005–14.
(38) Dijkstra A, De Vries H, Roijackers J. Long-term effectiveness of computer-generated tailored feedback in smoking cessation. Health Educ Res 1998; 13:207–14.
(39) Drossaert CH, Boer H, Seydel ER. Health education to improve repeat participation in the Dutch breast cancer screening programme: evaluation of a leaflet tailored to previous participants. Patient Educ Couns 1996;28: 121–31.
(40) Meldrum P, Turnbull D, Dobson HM, Colquhoun C, Gilmour WH, McIlwaine GM. Tailored written invitations for second round breast cancer screening: a randomised controlled trial. J Med Screen 1994;1:245–8.
(41) O'Connor AM, Tugwell P, Wells GA, Elmislie T, Jolly E, Hollingworth G, et al. Randomized trial of a portable, self-administered decision aid for postmenopausal women considering long-term preventive hormone therapy. Med Decis Making 1998;18:295–303.
(42) Kreuter MW, Strecher VJ. Do tailored behavior change messages enhance the effectiveness of health risk appraisal? Results from a randomized trial. Health Educ Res 1996;11:97–105.
(43) Dijkstra A, De Vries H, Roijackers J. Computerized tailored feedback to change cognitive determinants of smoking: a Dutch field experiment. Health Educ Res 1998;13:197–206.
(44) Dijkstra A, De Vries H, Roijackers J, van Breukelen G. Tailoring information to enhance quitting in smokers with low motivation to quit: three basic efficacy questions. Health Psychol 1998;17:513–9.
(45) Kreuter MW, Strecher VJ. Changing inaccurate perceptions of health risk: results from a randomized trial. Health Psychol 1995;14:56–63.
(46) Wroe AL, Salkovskis PM, Rimes KA. The prospect of predictive testing for personal risk: attitude and decision making. Behav Res Ther 1998;36: 599–619.
(47) Lloyd S, Watson M, Waites B, Meyer L, Eeles R, Ebbs S, et al. Familial breast cancer: a controlled study of risk perception, psychological morbidity and health beliefs in women attending for genetic counselling. Br J Cancer 1996;74:482–7.
(48) Lerman C, Narod S, Schulman K, Hughes C, Gomez-Caminero A, Bonney G, et al. BRCA1 testing in families with hereditary breast–ovarian cancer. A prospective study of patient decision making and outcomes. JAMA 1996; 275:1885–92.
(49) Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995;87:286–92.
(50) Watson M, Duvivier V, Wade Walsh M, Ashley S, Davidson J, Papaikonomou V, et al. Family history of breast cancer: what do women understand and recall about their genetic risk? J Med Genet 1998;35:731–8.
(51) Hallowell N, Statham H, Murton F, Green J, Richards M. Talking about chance: the presentation of risk information during genetic counseling for breast and ovarian cancer. J Genet Couns 1997;6:269–86.
(52) Welch HG, Burke W. Uncertainties in genetic testing for chronic disease. JAMA 1998;280:1525–7.
(53) Wagner TH. The effectiveness of mailed patient reminders on mammography screening: a meta-analysis. Am J Prev Med 1998;14:64–70.
(54) Fischhoff B. Why risk communication can be so difficult. Monogr Natl Cancer Inst 1999;25:7–13.
(55) Weinstein ND, Sandman PM. Some criteria for evaluating risk messages. Risk Anal 1993;13:103–14.
(56) de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999;36:99–105.
(57) Institute of Medicine and National Academy of Sciences. Exposure of the American people to iodine-131 from Nevada nuclear-bomb tests: review of the National Cancer Institute report and public health implications. Washington (DC): National Academy Press; 1998.
(58) Chamberlain M. New technologies in health communication. Am Behav Sci 1994;38:271–84.
(59) McCaul KD, Tulloch HE. Cancer screening decisions. Monogr Natl Cancer Inst 1999;25:52–8.
(60) Campbell MK, Bernhardt JM, Waldmiller M, Jackson B, Potenziani D, Weathers B, et al. Varying the message source in computer-tailored nutrition education. Patient Educ Couns 1999;36:157–69.

Notes

Supported by Public Health Service grants 1P01CA72099 and 1P50CA68438-02 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

We thank Tina Felix and Arline Sanchez for word processing as well as Scientific Consulting Group for editorial support.

The Visual Communication of Risk

Isaac M. Lipkus, J. G. Hollands

This paper 1) provides reasons why graphics should be effective aids to communicate risk; 2) reviews the use of visuals, especially graphical displays, to communicate risk; 3) discusses issues to consider when designing graphs to communicate risk; and 4) provides suggestions for future research. Key articles and materials were obtained from MEDLINE® and PsychInfo® databases, from reference article citations, and from discussion with experts in risk communication. Research has been devoted primarily to communicating risk magnitudes. Among the various graphical displays, the risk ladder appears to be a promising tool for communicating absolute and relative risks. Preliminary evidence suggests that people understand risk information presented in histograms and pie charts. Areas that need further attention include 1) applying theoretical models to the visual communication of risk, 2) testing which graphical displays can be applied best to different risk communication tasks (e.g., which graphs best convey absolute or relative risks), 3) communicating risk uncertainty, and 4) testing whether the lay public's perceptions and understanding of risk varies by graphical format and whether the addition of graphical displays improves comprehension substantially beyond numerical or narrative translations of risk and, if so, by how much. There is a need to ascertain the extent to which graphics and other visuals enhance the public's understanding of disease risk to facilitate decision-making and behavioral change processes. Nine suggestions are provided to help achieve these ends. [Monogr Natl Cancer Inst 1999; 25:149–63]

Communicating risk to the public poses formidable challenges *(1–4)*. Chief among them is helping the public understand risk, especially low-probability events (e.g., probabilities <.1). In most situations, risk is conveyed numerically (e.g., percentages or probabilities), and the literature is increasing on how different numerical expressions affect perceived risk (i.e., how people view risk) and decision-making processes [e.g., *(5–10)*]. However, visual displays, such as graphs and pictures, can be used in lieu of numbers or as adjuncts to aid in the further understanding of numerical risks. Unfortunately, little is known about how visual displays of risk independently, or in combination with numerical or narrative translations, affect perceived risk, decision-making processes, and, ultimately, behavior. Understanding how visual displays of risk affect these processes will become increasingly important as communication aids, especially with the accumulating evidence that a significant proportion of people have difficulty grasping and using numerical risk [i.e., are innumerate *(11–13)*].

This review has several objectives. First, we discuss how graphical displays can improve the communication of risk. We focus primarily on graphical displays because they are commonly used to transmit risk information. [Results pertaining to pictorials and other visual displays (e.g., posters or media) will be discussed when relevant.] Second, we discuss criteria to evaluate the effectiveness of graphical displays for communicating risk. Third, we summarize how graphical displays of risk affect risk perception, most notably risk magnitude. To narrow the focus, we review studies that experimentally tested how the addition of graphical displays to numerical or verbal text affected risk perception and other outcomes (e.g., intentions) or that used focus groups or one-on-one interviews to evaluate graphical displays of risk. Then, on the basis of findings in visual and graphical perception, we highlight issues to be considered when designing graphical displays to communicate risk. Last, we provide recommendations to advance the state of the art in visually communicating risk.

How Visual Displays Can Enhance the Communication of Risk

Visual displays (e.g., graphics) have desirable properties that can enhance the understanding of (numerical) risk. We propose that graphics possess at least three desirable properties for communicating risk.[1]

First, graphics reveal data patterns that may go undetected otherwise *(14–17)*. For example, line graphs are excellent for conveying trends in data *(17–20)*, whereas pie charts and divided bar graphs are the best choice for depicting proportions *(18,21–23)*. Second, specific graph types may evoke automatically specific mathematical operations *(21,22)*. People use graphical schemas to assist in viewing and interpreting the numerical information depicted in graphs *(24,25)*. Given a particular task (e.g., comparing risks), certain graphs allow the observer to process more effectively information than when numbers are presented alone. Third, unlike numbers, graphs can attract and hold people's attention because they display information in concrete, visual terms. [For a review of how vividness effects may affect attentional processes, *see (26,27)*.]

Criteria to Evaluate the Effectiveness of Visual Displays for Communicating Risk

What standards should be used to determine whether a visual display is useful for communicating risk? When possible, visual displays of risk should be evaluated on the basis of the following seven criteria suggested by Weinstein and Sandman *(28)*. Their criteria are 1) comprehension (do people understand the message and suggestions contained in the message?), 2) acceptance (do people agree with the interpretation and recommended action?),

Affiliations of authors: I. M. Lipkus, Duke University Medical Center, Durham, NC; J. G. Hollands, Defence and Civil Institute of Environmental Medicine, Toronto, Canada.

Correspondence to: Isaac M. Lipkus, Ph.D., 905 W. Main St., Box 34, Durham, NC 27701 (e-mail: Lipku001@mc.duke.edu).

See "Notes" following "References."

© Oxford University Press

3) dose-response consistency (do perceptions of risk and intentions vary according to the magnitude of the risk, such that increasing levels of risk are perceived as such within a hazard?), 4) hazard-response consistency (do people facing a hazard that is higher in risk perceive the risk as greater, show greater readiness to take action, or both than people exposed to a hazard of lower risk?), 5) uniformity (do people with the same exposure level interpret and react similarly to the information?), 6) audience evaluation (do people find the communication to be clear and helpful?), and 7) direction of communication errors (when interpretations and action plans depart from those intended, is it because people overreact and underreact?). In addition, if visuals affect risk perceptions by making a hazard easier to imagine, then measures that capture the ease of visualizing the risk(s) should be incorporated.

Studies Examining the Visual Communication of Risk

In this section, we review studies that used graphical displays to affect risk perceptions and other outcomes (e.g., intentions), focusing on risk ladders, stick or facial displays, line graphs, dots or marbles, pie charts, and histograms [*see (29)* for other examples]. Examples of these displays are illustrated in Fig. 1.

To be useful, graphs must communicate different risk characteristics, such as 1) risk magnitude (i.e., how large or how small is the risk), 2) relative risk (i.e., comparing magnitude of two risks), 3) cumulative risk (i.e., observing trends over time), 4) uncertainty (e.g., estimating amount of variability and range of scores), and 5) interactions (e.g., synergy), among risk factors. We review how graphs have been used to communicate these

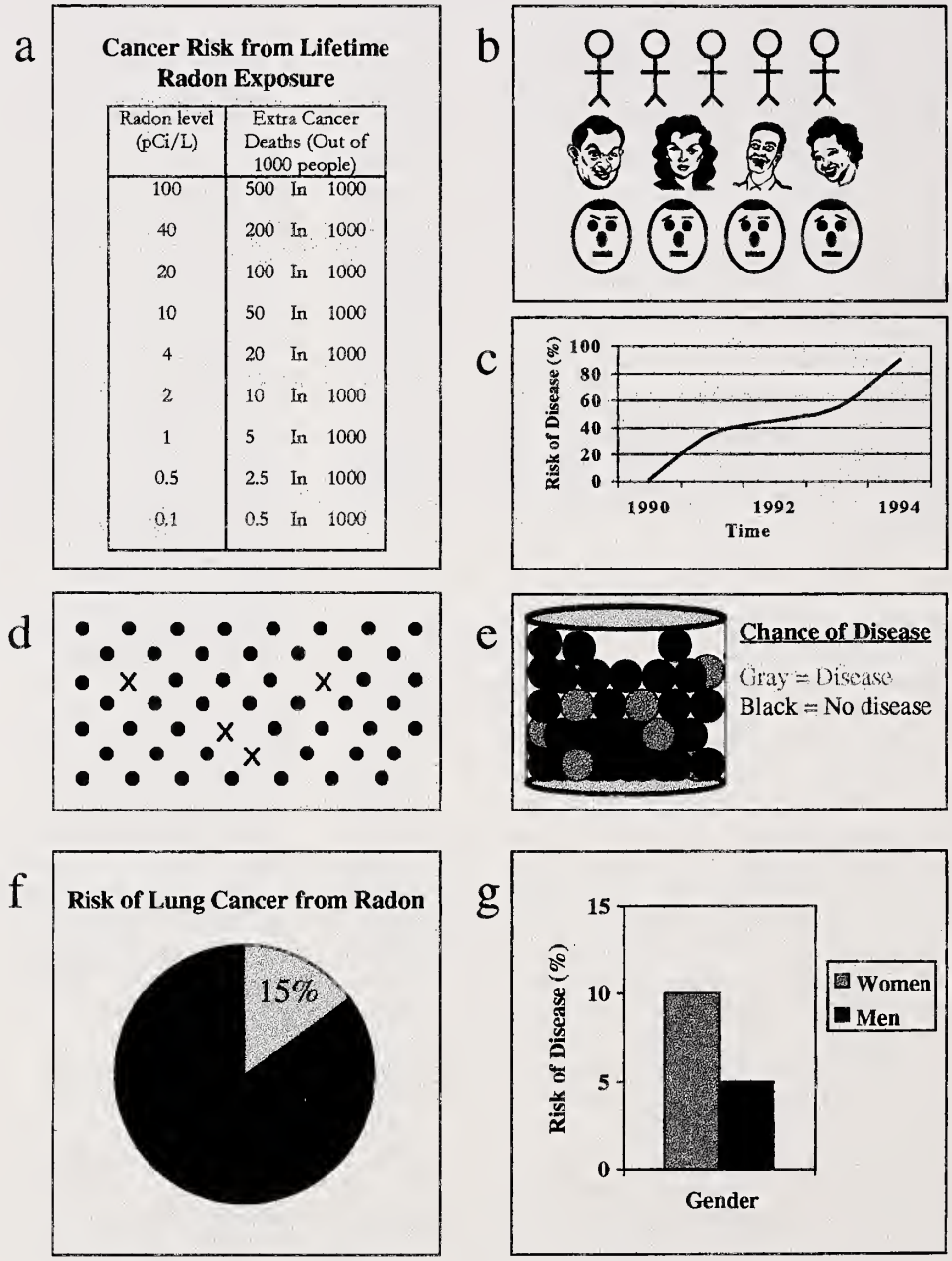


Fig. 1. Examples of visual displays that have been used to communicate risk. Researchers have used the following to illustrate risk: (*a*) risk ladder; (*b*) stick, human, and Chernoff faces; (*c*) line graph; (*d*) dots and Xs in which the Xs represent those affected by the hazard; (*e*) marbles; (*f*) pie chart (data are fictitious); and (*g*) histogram.

characteristics, beginning with the use of risk ladders. Because of the rarity of studies testing visual aids to communicate risk and of differences in design and outcomes among studies that tested visuals to eommunicate risk, detailed summaries pertaining to each graphical format are nearly impossible to compile. Rather, an overall summary of the extant literature is provided after the review.

Risk Ladder and Related Formats

The risk ladder has been used most extensively to describe environmental hazards [e.g., radon or asbestos *(30–38)*]. Typically, the risk ladder displays a range of risk magnitudes such that increasing risk is portrayed higher up on the ladder. Often, as a further aid, the risk in question is compared with other, perhaps more familiar, risks to encourage comprehension of small-probability events *(30,33–35,38)*. For example, the risk ladder used by Weinstein et al. *(33)*, and displayed in Fig. 2, equates different radon levels with quantities of cigarettes smoked and the extra number of cancer deaths. Therefore, the risk ladder communicates both risk magnitude and relative risk. In addition, the ladder can include other information, such as an action standard (i.e., a level at which a hazard poses a threat to be acted on, e.g., 4 pCi/L) and advice specifying how to help interpret the risk and what action to take, if any (*see* Fig. 2).

Extensive experimental research testing the utility of the risk ladder has been performed by Weinstein and colleagues who examined the communication of asbestos and radon risk [*(33,34,36,39); see (40)* for summary report]. There also are focus group impressions of the risk ladder *(41)*. In a series of studies *(33,34,36)*, Weinstein and colleagues presented homeowners with brochures with hypothetical test results of radon and asbestos that used different formats. For example, formats varied with respect to 1) location of risk on the ladder (i.e., radon or asbestos risk was high or low on the ladder), 2) the addition of a comparative risk (i.e., smoking; *see* Fig. 2), whether an action standard or advice (e.g., how soon you should act to remove radon) was presented,[2] 3) magnitudes of risks, and so forth. Although several outcomes were measured across studies (e.g., clarity or helpfulness of material and acceptance of action recommendations), we highlight two: perceptions of threat (composite measure of perceived risk, seriousness of level of risk, and concern and fear) and mitigation intentions (e.g., willingness to spend $1000 to reduce risk of radon or asbestos to 0; intention to take action).

There are four relevant findings. First, people's perceptions of threat were influenced by location of the risk on the ladder; people saw threat as greater at the top of the ladder. If the goal is to affect perceived risk, placing a risk closer to the top of ladder should increase people's perceived risk. Second, adding a risk ladder to a written action standard reduced perceived threat and mitigation intentions relative to the use of an action standard alone. Hence, if the goal is to lessen panic, a risk ladder may achieve this goal by allowing people to judge where they stand along a range of risk. People can compare other levels with their own. Third, the addition of a risk comparison (i.e., quantities of cigarettes smoked) did not affect perceived threat, mitigation intentions, or help people distinguish what is a high or low risk beyond that provided by an action standard or advice. However, adding a risk comparison made people feel that the information was more helpful and that they understood their risk better. Thus, trying to improve people's understanding of unfamiliar risks by comparing them with familiar risks may not strengthen the relationship between a person's actual risk and responses to that risk, although people may view the inclusion of the information as useful. Fourth, as shown in Fig. 2, having a risk ladder that includes risk probabilities (e.g., number of additional deaths because of radon or asbestos at different exposures), a risk comparison (i.e., cigarettes), an action standard (i.e., 4 pCi/L), and advice seems best to help people 1) distinguish between different risk levels for a single hazard, 2) develop appropriate mitigation intentions in accordance with their level of risk (i.e., people's intentions to pay for or take action to reduce a risk correspond with the actual level of risk such that mitigation intentions are greater at higher risks and vice versa), and 3) feel confident that they understand the risk.

In a study by Johnson and Slovic [*(42)*, study 2], college students were provided with information about the additional lifetime risk of getting cancer from a fictitious water contaminant presented as written text with or without a graph. The graph was a vertical line conveying cancer risk magnitudes of 0, 1 in 1000, to 1 in 100 at the bottom, middle, and top of the graph, respectively. Therefore, their risk ladder visually depicted a range of risks from low to high (0 to 1/100). The main outcomes were whether the addition of a graph affected how well people 1) noticed a range of risk estimates (e.g., are more apt to acknowledge that a risk can have different values ranging from low to high) assessed on a 4-point scale from seeing no risk to a great range of risk, 2) perceived trust in governmental agencies, and 3) perceived risk. Overall, the addition of a graph 1) increased the perceived *range* of risk (i.e., from no to great range) but did not affect perceived risk, and 2) decreased the perceived trustworthiness of the information. In addition, focus groups (study 3) reacted positively toward the visual portrayal of risk, claiming that the graph made the story clearer and more salient.

In sum, the risk ladder effectively helps people "anchor" a risk to upper- and lower-bound references points. Perceived risk is influenced by the location of risk perhaps more so than the actual numbers *(34,36)*. The efficacy of the risk ladder (e.g., to promote behavior change, understand one's risk, etc.) can be enhanced by the addition of an action standard and advice relevant to different risk levels *(33,35,37)*. Action standards and advice may influence significantly whether any actions to advert the risk are taken *(37)*. However, questions about its use remain *(40)*. For example, should the scales be logarithmic versus linear? What should the range of risk values be?

Stick and Facial Figures

Stick and facial figures have been used most extensively to aid relative risk judgments. In in-depth analyses of visual displays of risk, Stone et al. *(9)* examined how well stick and other visuals (bar graph or asterisks) communicated low-probability events (e.g., tire blowouts and serious gum disease).

Three studies were performed. In study 1, participants were shown the risks of a tire blowout using standard or improved tires in which the risk was reduced by half (i.e., 30 versus 15 serious injuries per 5 000 000 drivers for standard and improved, respectively). The risks were presented numerically or with added stick figures depicting the number injured (e.g., 15 or 30). A similar presentation was made for a standard and improved toothpaste. In study 2, participants were randomly assigned to

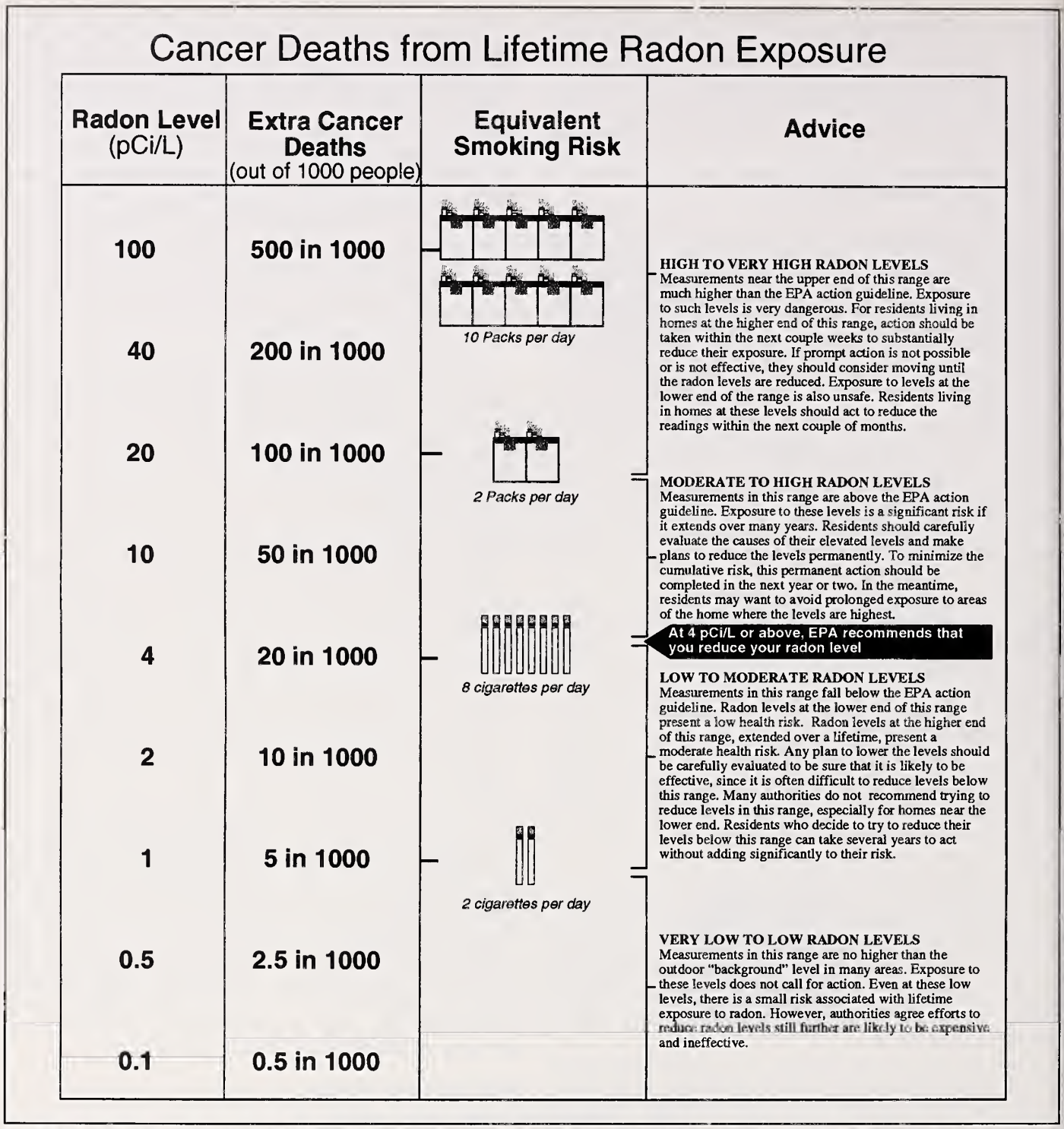


Fig. 2. Example of a risk ladder conveying the risks of radon. Radon levels are being compared with the number of cigarettes smoked and the number of extra cancer deaths. **On the right,** the ladder displays an action standard (pointing arrow of 4 pCi/L), along with advice on how to interpret radon levels and the action that is required, if any. Reprinted with permission of author *(33)*.

one of four conditions: 1) risk magnitude presented as numbers, 2) risk magnitude presented as numbers with stick figures—replicating the two experimental conditions in study 1, 3) replacing stick figures with asterisks, or 4) replacing stick figures with a bar graph. In study 3, participants were presented with stick figures or human faces. In each study, the main outcome was willingness to pay for the product (i.e., improved tires or toothpaste).

Overall, participants were more willing to pay for an improved product when graphics (stick figures, asterisk, or bar graph) were added to the numerical presentation (studies 1 and 2, respectively). That is, participants became more risk averse with the addition of a graphical figure. Willingness to pay was not affected by the kind of displays examined in studies 2 and 3. In sum, the studies by Stone et al. suggest that graphical displays of comparative risk increase risk aversion relative to presenting

numbers alone. However, visual displays did not produce greater risk aversion for higher-probability events [e.g., .40 *(43)*].

Stephenson and Witte *(44)* tested whether vivid facial displays can promote skin protective behaviors through fear appeals related to skin cancer. College students randomly received a text message alone or with accompanying visuals (e.g., pictures of people with advanced skin cancer)—how effective sun block was at reducing skin cancer was also manipulated. Participants in the latter condition expressed greater intent to take preventive measures to protect their skin, but there were no differences in perceptions of fear and threat (i.e., risk by severity interaction) between the text and the text with visuals. Thus, the use of high threat visuals may not enhance the effectiveness of fear appeals, as the literature on vividness effects would suggest [for reviews, *see (26,27)*], but behavioral intentions can be modified without necessarily evoking fear.

Stick and facial figures have also been used to illustrate how many women, out of a certain number, are expected to get breast cancer. Such displays have been used to show the relationship between breast cancer risk and age and the effects of tamoxifen and hormone replacement therapy. This approach assumes that showing relative frequencies will maximize women's understanding of (low magnitude) risks by focusing, for example, on the numerator (showing the number of women affected) versus the denominator (how many women in the background are not affected).

There is a lack of research testing the effectiveness of communicating breast cancer risk using stick figures or women's faces compared with other visuals.[3] Pilot data obtained from the Duke Risk Communication Laboratory suggest that women may not prefer the former mode of communicating risk. Participants were presented with 100 female figures on a 8½ × 11 page. Figures with breast cancer had "X"s across them. Women thought the picture was too complex, or "busy," and did not like the idea of having women who were affected "X"ed out. However, results from the study by Stone et al. *(9)* discussed earlier would suggest that presenting breast cancer risk via the use of stick figures or faces versus numbers only would induce risk aversion, which may, in turn, increase the likelihood of breast cancer screening or genetic testing.

In sum, similar to other visuals (e.g., asterisks or histograms), stick figures and facial displays may affect perceived risk for smaller probabilities, leading people to be risk averse. Preliminary evidence suggests that people do not desire these displays to convey the risks and consequences of events like cancer. Thus, stick figures or facial displays may affect behavioral change despite the unappealing nature of the risk information.

Line Graphs

Line graphs are effective for communicating trends in data *(17–20)*. Explicit tests of whether the addition of a line graph to numerical or written text improves people's understanding of risk over time (e.g., cumulative risk) are unavailable. However, Mazur and Hickman *(45)* tested how physicians', medical students', and patients' treatment preferences for a serious yet unidentified medical condition were influenced by three 5-year survival curves that differed with respect to the amount of area under the curve. Overall, unlike physicians and medical students, patients' treatment preferences were most influenced by the location of the immediate (year 0) and end point (year 5) data. Medical students' and physicians' treatment preference was also influenced by the shape of the curve pertaining to years 2 to 4.

Dots and Related Formats

A few experimental studies have tested the efficacy of using a field of dots to communicate different probabilities of disease.[4] Kaplan et al. *(46)* informed 240 college undergraduates of one of three probabilities of having a negative reaction (i.e., nerve damage) to a flu vaccine (Guillian-Barré syndrome). The probabilities were 1/1000, 1/10 000, and 1/100 000. Half the participants were provided with a probability only; the other half were provided with a visual display of dots representing one of the three risks. Visually, those who reviewed a risk of 1/1000 saw 1/10th of a page covered with 1000 dots, those who reviewed a risk of 1/10 000 saw a page covered with 10 000 dots, and those who reviewed a risk of 1/100 000 saw 10 pages covered with 10 000 dots per page. Participants were asked how likely they were to get vaccinated. Regardless of the probability of side effects, the visual display increased the likelihood that participants would take the vaccine.

Weinstein et al. *(47)* performed an extension and attempted to replicate the results of Kaplan et al. In study 1, college students were presented with two decisions. The first decision was whether to get vaccinated against the flu given the risk of possible nerve damage—thus replicating the influenza scenario of Kaplan et al. The second decision was whether to move out of a dormitory to avoid the risk of minor nerve damage from improperly applied pesticides. Two levels of risk for side effects were used: 1 in 50 or 1 in 10 000. These risks were presented as odds, odds plus a grid of dots to represent the denominator of the odds ratio (i.e., total population), and a condition whereby an "X" was included among the dots to represent the numerator (the number affected) of the odds ratio (*see,* for example, Fig. 1, d). In addition, to more powerfully test whether participants are more influenced by the numerator or denominator of the odds ratio, a third condition was added involving a risk of 200 in 10 000, represented using each of the three formats described above. Format did not affect intentions or perceived threat. In study 2, four formats illustrating the risk of having a side effect from a flu vaccine (1 in 10 000) were used: 1) odds only; 2) odds plus a grid of 10 000 dots; 3) odds plus a grid of 10 000 dots plus a single dot on a separate page; and 4) a format identical to format 2, except that the dilemma, format, and questionnaire assessing risk and intention to act were presented one page at a time. Format did not affect intention; however, format 3 elicited more perceived threat than format 4. Overall, in neither study did any visual display differentially affect threat perception and intention compared with the odds-only format.

Similar to dots, marbles have been used to convey the risk of getting breast cancer (*see,* for example, Fig. 1, e). Baty et al. *(48)* presented women who had the BRCA1 mutation with the cumulative risk (i.e., aggregated risk over time) of getting breast cancer and ovarian cancer by age, using different colored marbles randomly distributed in a jar. These women also were presented with the cumulative risks of getting breast cancer and ovarian cancer for women without the mutation, again using marbles in a jar. Baty et al. reported that these displays helped several women understand the effects of carrier status on getting breast cancer and ovarian cancer. However, they pointed out that some women may be disturbed by this presentational style, in comparison with tabular information.

Pie Charts

Pie charts are effective for conveying proportions *(18,21–23)*. Fig. 3 shows a visual developed by the National Cancer Institute that depicts the risk of getting lung cancer as a function of smoking and radon exposure *(49)*.[5] Several participants in focus groups found the visual confusing, because they lacked knowledge about radon. In addition, although participants understood that the graph showed several risk factors linked to smoking and that smoking was a greater risk than radon exposure, some felt it was too complex (i.e., its main idea was not clear). Furthermore, a study by Hampson et al. *(50)* showed that the use of three pie charts failed to help communicate the synergistic (i.e., interactive) effects of smoking and radon exposure on lung cancer risk. Participants were shown a slice of a pie, specifying the total risk of getting lung cancer from radon exposure and smoking, that was *greater* than the sum of the pie slices for each risk factor displayed separately. These results suggest that 1) relatively straightforward visuals are easily misinterpreted and 2) visuals should be accompanied by brief captions that state the take-home message(s), especially for unfamiliar risks.

Devousges and colleagues *(31,41)* used focus groups to pilot test the effectiveness of using several pie charts—which they called probability or risk circles—to portray environmental risks. Participants were presented with two probability circles. One portrayed the risk of exposure to an environmental hazard, and the second portrayed the risk that exposure would produce detrimental effects (*see* top two circles in Fig. 4). Therefore, participants had to first imagine being exposed to an environmental hazard (i.e., landing on the pie slice that represented being exposed) and then similarly imagining the risks of harm if exposed. Through group discussions, participants expressed their difficulties understanding 1) how the shaded part of the pie conveyed chance (i.e., risk) and 2) exposure and effect as separate events or effect being contingent on first being exposed. Consequently, a third pie chart was added that displayed combined risk (i.e., exposure and effect; *see* bottom of Fig. 4). This third pie chart aided their understanding of the distinct probabilities. Therefore, a pie chart showing joint probabilities is more effective than two pie charts, each showing single-risk probabilities.

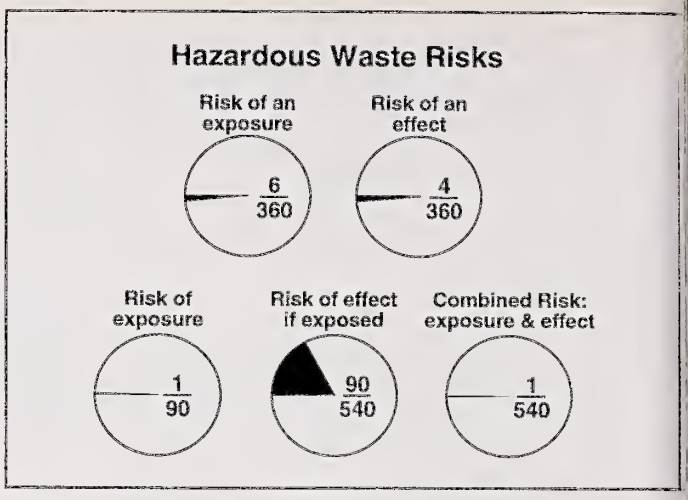


Fig. 4. Examples of risk circles, or probability circles, to illustrate risk of environmental hazard. The **top two circles** were less understood than the **bottom three circles**. The addition of the circle that combined the probabilities of exposure and effect aided comprehension. Reprinted with permission from the author *(31)*.

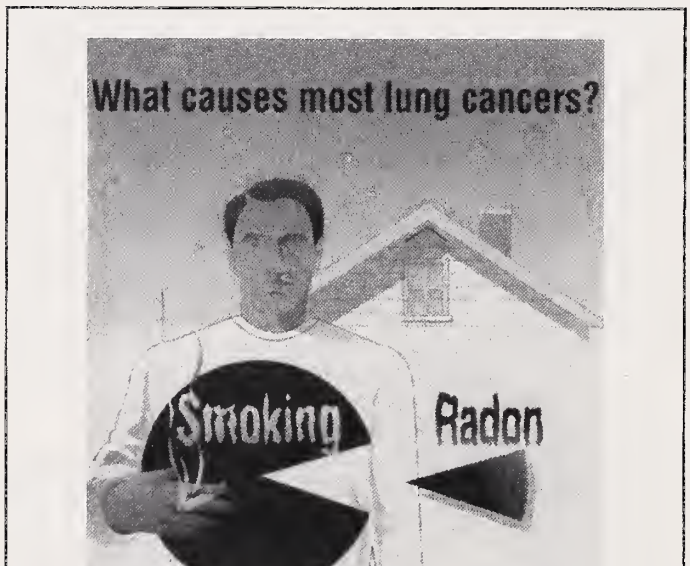


Fig. 3. Pie chart developed by the National Cancer Institute and evaluated by focus groups to depict lung cancer risk as a function of smoking and radon exposure. Reprinted with permission from the National Cancer Institute *(49)*.

Histograms

Histograms have been used to convey smoking-related diseases in cessation manuals, such as the American Cancer Society's *Pathways to Freedom: Winning the War Against Tobacco*. This cessation manual portrayed the leading causes of death for African-Americans (i.e., homicide, car crashes, smoking-related diseases, drug abuse, and acquired immunodeficiency syndrome) in a histogram. Pilot testing of this graph at the Duke Risk Communication Lab showed that all participants understood that smoking was the leading cause of death for African-Americans and were quite alarmed by this fact.

The National Cancer Institute *(49)* used focus groups to test reactions toward a histogram to convey breast cancer risk (*see* Fig. 5). The histogram plotted the number of cases of breast cancer for different age categories (e.g., 40–49 years, etc.). Overall, the histogram was well received and understood. Women appreciated the gray shading to highlight the age groups at greatest risk. However, some participants were disturbed by the numbers and questioned their validity (e.g., why were the statistics of the number of women affected by breast cancer not more up to date?).

Although the research linking histograms with perceptions of risk is sparse, it appears that people readily understand and find histograms helpful and that they may induce risk aversion compared with numbers alone *(9)*. Whether risk perception is more affected by numerical risks or the height of the bars *(33)* remains to be tested—this can be determined by showing risk data using other graphical formats as well as the histogram.

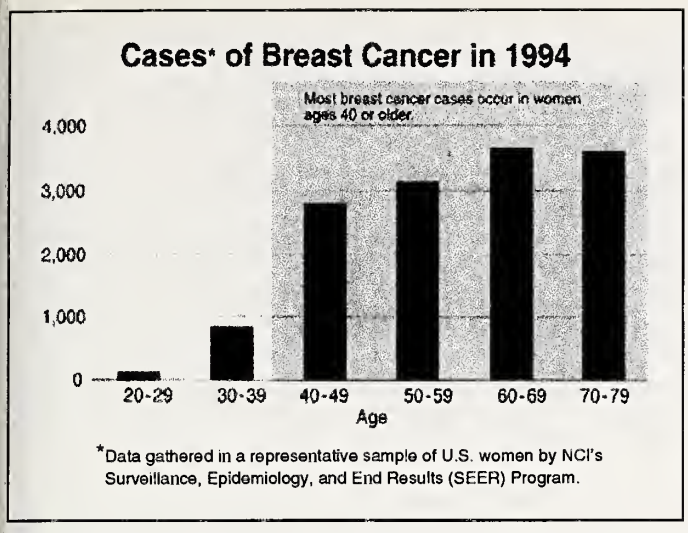


Fig. 5. Histogram developed by the National Cancer Institute and evaluated by focus groups to depict breast cancer risk as a function of age. Reprinted with permission from the National Cancer Institute *(49)*.

Miscellaneous Visuals

A few studies have assessed how variability in data (i.e., distribution of data points) is perceived. Although these data do not address risk perceptions per se, they provide some insights into how people respond to variability in data. We review these few studies below.

Lathrop *(51)* replicated and extended research by Hoffstatter [cited in *(51)*], who found that, when people were asked to judge the variability associated with a bundle of sticks, they appeared to weigh the standard deviation by the reciprocal of the mean. That is, as values increase, perceived variability decreases. Lathrop showed cards depicting a set of lines to subjects and varied the order of presentation. Some cards had lines that alternated radically between long and short, whereas others did not. He found that the sequencing of the lines affected variability estimates, even when subjects were told to ignore the sequence. These results suggest that, if risk data are presented visually in different sequences, they may affect perceived variability, which can subsequently affect subjective probabilities and how these probabilities are combined (e.g., greater variability may lead to inflated probability estimates).

Legge et al. *(52)* examined how observers detect differences in variances between pairs of datasets sampled from Gaussian distributions. These data were displayed in three formats: tables, scatterplots, and luminance-coded displays. Performance was best for scatterplots and worst for tables.

Ibrekk and Morgan *(53)* examined how educated and semi-technical participants used nine visual formats showing quantitative distributions. The nine displays included an error bar, pie charts, histograms, conventional probability density plots, Tukey Box, and a cumulative probability plot. For the first of a two-part questionnaire study, the displays depicted meteorological information about snowfall. Participants were shown the nine displays and asked to indicate for each display 1) the best estimate of how much snow would fall, 2) the chance that more than 2 inches of snow would fall, and 3) the chance that the amount of snowfall would be between 2 and 12 inches. In part 2, participants were presented with these displays and asked similar questions to predict amount of water depth from a flood. In part 2, the questionnaire also contained a series of nontechnical explanations of the meaning and use of each of the displays in the context of predicting water depth from a flood.

When the task was to convey a single best estimate (e.g., amount of snowfall), graphs that explicitly portrayed the mean produced the most accurate responses (i.e., error bar and modified box plot; *see* Fig. 6, a and b). When the task was to estimate the probability that a quantity exceeded a certain amount (e.g., probability of snowfall greater than 2 inches) or fell within certain amounts (e.g., probability of more than 2 inches but less than 12 inches of snow), performance was best for estimating probabilities using a plot with a cumulative distribution function (CDF) (*see* Fig. 6, c) after explaining how it should be used—the pie chart was superior to the CDF but only before explanation. Ibrekk and Morgan *(53)* speculated that the combined use of cumulative probability and density plots would be most effective for communicating uncertain quantities (*see* Fig. 7), although this idea has not been explicitly tested.

Summary of Visual Displays of Communicating Risk

Despite the urging of several risk communication researchers [e.g., *(1,54,55)*], little research has tested the efficacy of visual formats for communicating risk per se. In this section, we summarize the research of the studies reviewed above.

Visuals May Help

Most risk information is portrayed as numbers alone (e.g., in tables) or as numbers with narrative translations. Visuals are assumed to help individuals understand and summarize risk in-

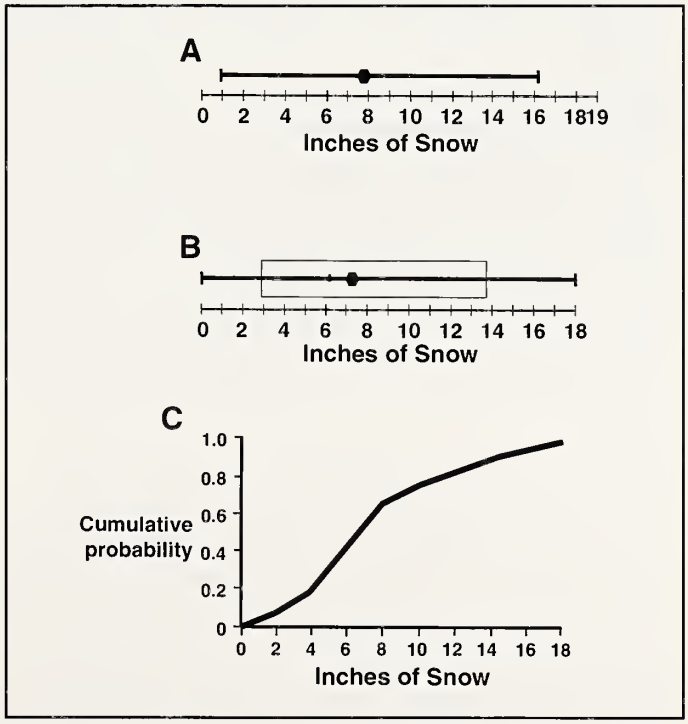


Fig. 6. Graphics that effectively convey point estimates and uncertainty data. **A)** error bar and **B)** modified box plot were most effective for communicating a single estimate (e.g., mean estimate of amount of snowfall), as indicated by the dot on the line. **C)** The cumulative probability plot was most effective for communicating the probability that an amount would exceed or be between certain values. Reprinted with permission from Risk Analysis *(53)*.

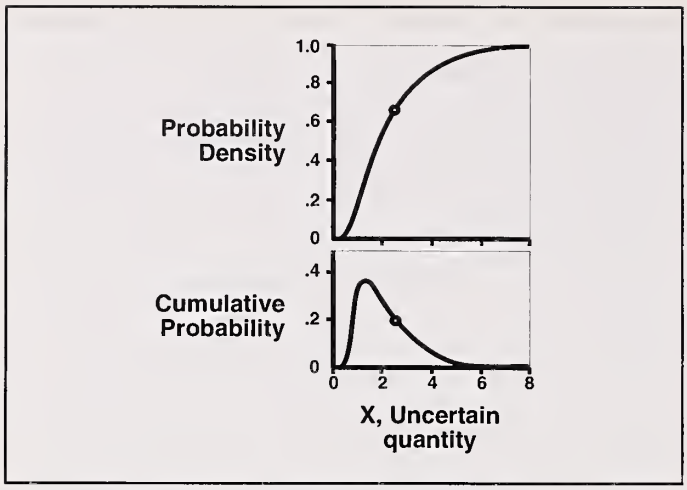


Fig. 7. Ibrekk and Morgan's recommended graphical plots to communicate quantitative uncertainties. This example of a cumulative distribution function is plotted directly below the probability density function with the same horizontal scale and with the location of the mean marked by a **dot.** Reprinted with permission from Risk Analysis *(53).*

formation. Most, but not all [e.g., *(47)*], of the preliminary *experimental* evidence suggests that combining visuals with numerical and written information does affect several outcomes, such as the perceived helpfulness of the information, perceived risk (e.g., threat), mitigation intentions, and trust in government officials *(9,35–37,39,42,46).* The tentative conclusion that visuals are useful for communicating risk is based on limited experimental studies that involved different hazards (e.g., radon or asbestos, product risks), risk magnitudes, type of risk being communicated (i.e., absolute versus relative risk), format of communicating risk (e.g., risk ladder, stick figures, or dots), and outcomes.

Despite Their Popularity, Experimental Tests of the Effectiveness of Visuals Are Few

The mass media (e.g., newspapers) and health communication researchers often use graphics to communicate risk, assuming that these displays will improve the public's understanding of risk and act accordingly. There is still little experimental research testing whether the lay public's perceptions and understanding of risk vary by graphical format and whether the addition of graphical displays improves comprehension significantly beyond numerical or narrative translations of risk and, if so, by how much.

Research Is Largely Atheoretical

With few exceptions *(9,31,33–35,39),* most research using graphs and other visuals specifically to communicate risk has been atheoretical. It is not clear why particular graphs or visuals are chosen. For research in this area to advance, the application of theory and findings from other fields (e.g., visual perceptions and psychophysics, human factors, or visual persuasion) may be useful. Nonetheless, hypotheses have been put forth with respect to 1) how stick and other visuals may induce risk aversion (e.g., willingness to pay) when added to numerical risk information *(9);* 2) how the placement of risk information on a risk ladder affects perceived risk and intentions *(33,36,37);* 3) how an action standard and advice, when added to a risk ladder, affect perceived risk, understanding, intentions, and behaviors *(34–37);* and 4) how visually distinguishing the numerator from the denominator of an odds ratio affects perceived risk and intentions when added to numerical information *(46,47).*

Overall, it is still unclear why the addition of a visual display to numerical risk information would induce greater risk aversion *(9).* Similarly, results are inconclusive whether visually emphasizing the numerator and denominator of an odds ratio (e.g. showing "X"s to represent number affected versus dots to represent total population) affects perceived risk and intentions. People are sensitive to placement of risks on the risk ladder. Furthermore, as discussed previously, adding an action standard and advice affects perceived risk, intentions, and possibly behavior. Nonetheless, a comprehensive theoretical model linking visuals to risk communication that can incorporate these findings is needed.

Impact of Task Is Ignored

The use of particular visuals has not been linked explicitly to specific risk communication tasks. The most extensive research examining the effectiveness of visuals has examined the use of the risk ladder and the testing of stick and facial figures, primarily for the communication of absolute and relative risks. What is known is that 1) risk ladders are effective at conveying magnitude and range of risk, based on the positioning of the hazard(s) on the ladder; 2) stick figures, faces, asterisks, and histograms are roughly equivalent in their ability to induce risk aversion when added to numbers, although it is unclear whether people's understanding of relative risk is improved comparing visuals displays with numbers only *(9);* and 3) people have difficulty understanding low-probability events, even with the aid of visuals.

Research that examines which graphics and visuals are best matched to particular risk communication tasks is needed. The graphical perception literature shows that certain graphs are well suited for particular tasks (e.g., evaluating trends in risk over time, judging proportions, etc.). We discuss this research and its applications to risk communication in a later section.

Few Visuals Are Known to Be Effective at Communicating Uncertainty

While a few attempts have been made to visually communicate the quantitative uncertainties of environmental health risks, we have yet to begin testing visuals for communicating uncertainties about cancer risk per se. What is known is that placing the upper and lower bound of risk on a vertical scale, akin to a risk ladder, helps people understand that the risk of an adverse outcome can take on high and low values (i.e., see a range of risk) and that scatterplots are effective for conveying variability (i.e., dispersion) in data. In addition, if a series of data is presented whereby the ordering of the magnitudes differs substantially (e.g., have an alternating sequence of presenting low and high risk), people's perceptions of variability and perhaps risk may be affected. However, it is not clear how perceived variability would affect perceived risk. For example, if variability is perceived as uncertainty or instability, people may perceive the information as less trustworthy and consequently be less inclined to alter their risk perception and actions. If the goal is to communicate probabilities for a range of outcomes (e.g., probability that we will get rain in the amount of 1 to 4 inches), the combined cumulative and probability density plots suggested by

Ibrekk and Morgan [*(53); see* Fig. 7] might prove useful after explaining to people how to use them.

Few Standards Used to Evaluate Visual Representations of Risk

To date, visuals have been evaluated on the basis of their ability to affect primarily people's perceptions of risk magnitude and relative risk and intentions to modify risk-related behaviors (e.g., test for radon). Little is known about how visuals affect other outcomes as part of the risk communication processes, as discussed in the section on criteria used to evaluate the effectiveness of visual displays to communicate risk. The work by Weinstein and colleagues *(33,34,36),* pertaining to communicating risks of exposure to radon or asbestos, ideally exemplifies the explicit use of these criteria.

Adding Reference Points to Graphs Affect Risk Perceptions and Intentions

Graphs that contain a reference point (e.g., action standard, colors that highlight level of risk, or arrow showing high risk) to indicate level of hazard threat (i.e., low, moderate, or high risk) affect risk perceptions, intentions, and possibly behaviors *(33–35,37).* When possible, graphs should contain a reference point indicating when a hazard reaches a level requiring action (i.e., an action standard) along with advice about what action to take at that level. Information also should be included about how to interpret risk below and above the action standard and what, if any, appropriate action is needed at these levels. Indeed, it is relatively easy to incorporate a reference point and some descriptions into some graphs, and different graphical formats make it easier to compare a particular quantity with a reference point *(56).*

Guidelines for Maximizing the Effectiveness of Graphs

In the previous sections, we provided general conclusions about the use of graphical displays but provided no guidelines to maximize their effectiveness. In this section, we discuss seven guidelines pertinent to graphical depiction of risk data. [Readers interested in further guidelines for construction should consult Gillan et al. *(57).*]

No single graphical format will perform optimally in all situations. Rather, the effectiveness of a display will be affected by several factors, such as 1) the display characteristics (e.g., use of colors, width of lines, or type and spacing of legends), 2) conditions of presentation (e.g., lighting or time pressure), 3) data complexity (e.g., number of data points or configuration of the display), 4) the task (i.e., purpose, for example, to assess trends), 5) user characteristics (e.g., cognitive styles), and 6) the criterion for choosing the display (e.g., speed of performance or accuracy) *(19).* In our discussion, we assume that risk data can be plotted graphically similar to other data (e.g., finances, sports, or weather). Furthermore, we assume that what is known generally about graphical perception should apply specifically to the graphical depiction of risk data.

Avoid Areas or Volumes to Depict Quantities

People make errors when judging graphs. That is, people overestimate or underestimate quantities compared with their true values. These biases may be because of optical illusions [e.g., Poggendorf illusion *(58)*], distortion in memory (e.g., remembering more symmetry than exists [*see (59)* for review]), or perceptual limitations in estimating the magnitudes of certain kinds of stimuli (e.g., area and volume). The key point is that graph designers should avoid using graphical elements that lead to perceptual biases. We briefly review a few biases that are relevant to presenting risk data.

Individuals show biases in estimating physical magnitudes of objects. According to Stevens and Gallanter *(60)* and Stevens *(61),* the relationship between perceived and actual magnitude is described by the power function $p(x) = cx^{\beta}$, in which $p(x)$ is the perceived magnitude of the object, c is a constant, x is the actual magnitude, and $^{\beta}$ is an exponent. When the exponent is 1, there is no bias in the perceived magnitude; when the exponent is greater than 1, each incremental increase in the physical magnitude results in greater and greater estimates of perceived magnitude, producing a positively accelerated function (i.e., response expansion). When the exponent is less than 1, each incremental increase in physical magnitude results in smaller and smaller increases in perceived magnitude, producing a negatively accelerated function (response compression). In particular, response compression has been shown for estimates of areas and volume (dimensions often used to code variables in graphs), so that large areas and volume are underestimated. In contrast, lengths are usually perceived without bias. Illustrations of these biases are depicted in Fig. 8.

In addition, the biases described by Stevens's law appear to affect more complex judgments when multiple quantities are used, such as judgments of proportion (i.e., what proportion is *A* of *B*?). When using graphs depicting proportions (e.g., pie charts or stacked bar graphs), people show a cyclical bias pattern, such that proportions of 0–.25 and .50–.75 are overestimated and proportions of .25–.50 and .75–1 are underestimated [*see* Fig. 8, B–D, and *(56,62)* for review]. This bias can be reduced by adding reference points, such as tick marks, to the graph. For example, as shown in Fig. 8, D, placing a tick at 0, 90, 180, and 270 degrees of a pie chart has been shown to reduce errors in estimating proportions *(56).* Therefore, researchers who wish to illustrate proportional judgment of risk (e.g., proportion of breast cancer relative to all cancers) may reduce error by adding reference points to a pie chart. In sum, making errors in relative judgment can be reduced by 1) avoiding the use of stimuli whose Stevens's exponents differ from 1, and 2) increasing the use of reference points (e.g., tick marks) in the graph.

Consider the Task

What are graph readers being asked to do with the information depicted in the graph—to read the exact value of a risk estimate, compare the value of two risks, assess trends, or judge proportions? Consistent with Sparrow's *(63)* notion of task or display compatibility, a display should provide the relevant information such that it minimizes the amount of computational effort and transformations needed to perform these tasks.

To assist in this endeavor, Carswell *(64)* devised a taxonomy, incorporating the following four basic graphical tasks: 1) point reading, 2) local comparisons, 3) global comparisons, and 4) synthesis. In point reading, the observer judges the value of a single graphical element (e.g., what is the value of risk expressed in a histogram?). In local comparisons, the observer compares two graphical elements (e.g., which person, *A* or *B,* has a higher risk of getting breast cancer?). For global comparisons, the ob-

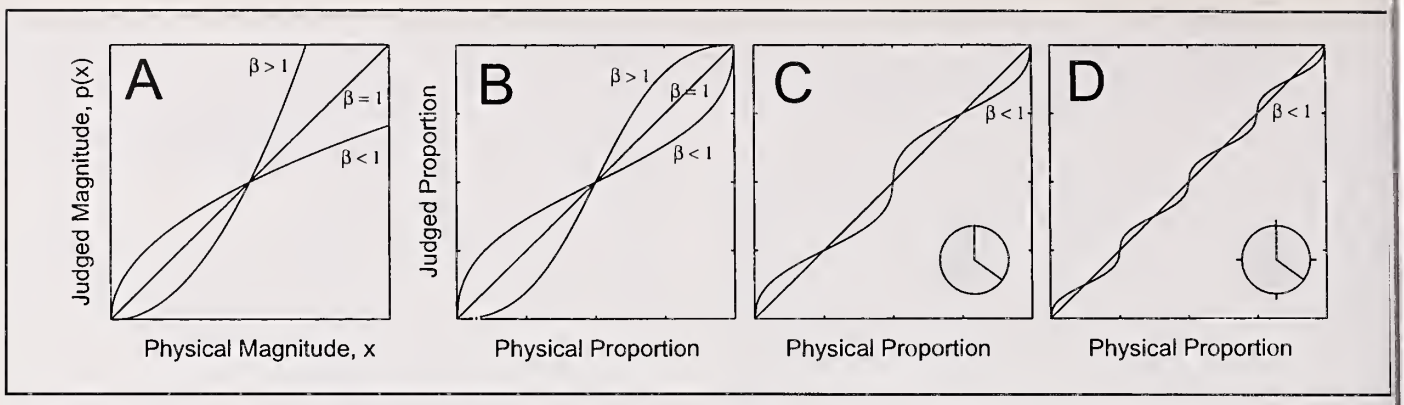


Fig. 8. Bias in perceptual judgment. **Panel A** shows the bias in single estimates of magnitude described by Stevens's law. β <1 when judging areas or volumes, β = 1 when judging lengths. **Panels B, C,** and **D** show the cyclical bias in proportion judgments that is because of the bias in the estimate of each single quantity. The patterns in **panels C** and **D** are seen when reference points at halves or quarters are used, respectively. The pattern in **panel C** is seen with common graph types, such as pie charts and divided (stacked) bar graphs. Error in judgment can be reduced by adding tick marks to a pie chart at the quarters, resulting in the pattern shown in **panel D.**

server compares quantities derived from other quantities shown in the graph (e.g., is person *A*'s risk for breast cancer less than that of *B*'s over two time periods?). For synthesis judgments, the observer needs to consider all data points to make a general, integrative judgment (e.g., is person *A*'s risk of breast cancer increasing or decreasing over time?).

Carswell's taxonomy can be applied to consumer risk information. A consumer might wish to extract the following types of information: 1) risk magnitude (i.e., how large or small the risk is), 2) relative risk (i.e., comparing the magnitude of two risks), 3) cumulative risk (i.e., observing trends over time), 4) uncertainty (e.g., estimating amount of variability or range of scores), or 5) interactions (e.g., synergy) among risk factors. Applying Carswell's taxonomy, the estimation of risk magnitude would be an example of point reading. Comparison of two risk magnitudes would constitute local comparisons, whereas assessing cumulative risk, uncertainty, and interactions among risk factors would most likely constitute global comparisons and synthesis judgments. Of import, point reading and local comparisons can be considered fairly low level, focused tasks, whereas global comparisons and synthesis judgments involve integrating several pieces of information (e.g., how variables are related) to see the "big picture." Some graphical displays are better suited for focused rather than integrated tasks, and we discuss these below.

When the task is to communicate a precise risk magnitude (i.e., point reading) or to compare two risks (i.e., local comparisons), two fairly low level focused tasks, the graph designers should make use of the ranking of elementary perceptual tasks as proposed by Cleveland and McGill *(16,65)*. According to Cleveland and McGill, individuals are least accurate at extracting information using 1) position along a common aligned scale (e.g., bar charts, histograms, or dot charts); 2) position on a common nonaligned scale (e.g., polygon displays with reference axes or scatterplots); 3) length (e.g., polygon displays without reference axes); 4) angle or slopes (e.g., pie charts, disks or meters); 5) area (circles or blobs); 6) volume, density, or color saturation (e.g., cubes); and 7) color hue (e.g., statistical maps with color coding). Therefore, if people are asked to evaluate the magnitude of a risk or asked to compare risks, line graphs, bar charts, histograms, and dot charts are likely to lead to greater accuracy in estimation than are other graphical formats. For example, the effectiveness of the risk ladder is probably because of its resemblance to a divided bar chart, which capitalizes on position along a common scale. If the task is to make proportional judgments of risk, a focused task that is served well using area to illustrate quantities, pie charts should be used *(18,21–23)*. Nightingale roses also may be effective for proportion judgments. Like pie charts, Nightingale roses are circular and segmented; however, they differ from pie charts in one important way. In a pie chart, each segment shares a common radius but has a central angle that varies as a function of the data. In a Nightingale rose, each segment subtends the same central angle, but the length of the segment differs as the square root of the radius *(66)* as shown in Fig. 9.

Overall, the ranking of elementary perceptual tasks by Cleveland and McGill *(16,65)* serves us well for focused risk communication tasks (e.g., point reading, comparing two risk magnitudes along a common scale, etc.). However, their rankings do poorly for addressing integrative tasks that require seeing the "bigger picture" *(64)*. A model for visual display design, called the proximity compatibility principle, makes predictions about which task–graph combinations should be most effective for focused and integrative tasks *(67)*.

According to the proximity compatibility principle, there are two important components to consider: display proximity and processing proximity. Display proximity relates to how similar graphical elements are to each other. Graphical elements will be

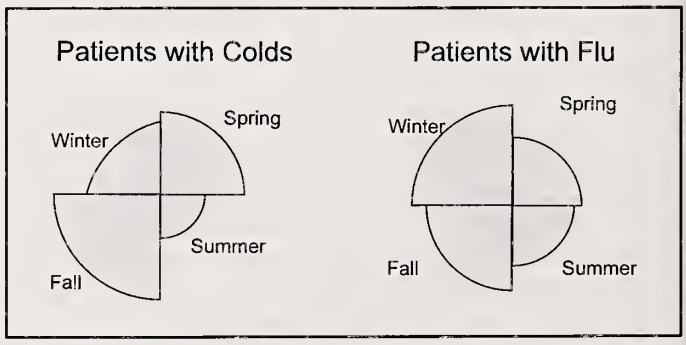


Fig. 9. Example of a Nightingale rose. For each rose, a circle is divided into multiple regions of equal angle; the radius of each slice is used to depict the quantity of interest. Because the data for each season are in the same position in each rose, it is easy to compare them. The data are fictional.

perceived as more similar to the extent that they are close together, share the same colors, use the same physical dimensions (e.g., length or height), or use the same code (e.g., both are digital or both are analog). Graphical elements that are viewed as similar and integrated have high display proximity. Processing proximity refers to whether two or more informational sources are used within the same task. Tasks that require integrating several pieces of information are said to require high processing proximity (e.g., estimate whether there is an increase or decrease in the slope of a scatterplot). Tasks that do not require the integration of several information sources are low in processing proximity (e.g., determine the value of a symbol in a graph). Hence, on the basis of these two dimensions, tasks that require high processing capacity (e.g., local comparison or synthesis tasks) should have high display proximity (e.g., integrate features of the graph to make them appear as one unit so that it is easier for the observer to use all the information, e.g., line graphs). Tasks that require low processing proximity (point reading or local comparisons) should have low display proximity (e.g., create graphs that highlight specific informational elements).

The proximity compatibility principle is in agreement with Cleveland and McGill's ranking system for focused tasks. However, the proximity compatibility principle makes different predictions than Cleveland and McGill's ranking system for integrative tasks, because these tasks are best preformed with integrated displays that often use area and volume to communicate quantities. Hence, for integrative risk communication tasks (e.g., global comparison or syntheses tasks), it is best to use displays that 1) highlight emergent features of the display, such as symmetry, area, and linearity, and 2) increase the spatial proximity among elements. For example, tasks that require seeing changes in several variables at once may benefit from facial displays in which different facial characteristics (e.g., eyes or mouth) convey different levels of multivariate data [e.g., Chernoff face displays *(68–70); see* Fig. 1, c, bottom row). Thus, facial displays may be useful for displaying synergy among risk factors, where specific features can be used to designate type and level of risks. In addition, integrative tasks are served well by such formats as line graphs *(71)*. Line graphs should be used when the main tasks are to assess trends in risk over time or to view the interaction between variables.

Minimize Number of Mental Operations

When a graph reader reviews a graph, he or she will perform a sequence of perceptual operations, cognitive operations, or both. A graph designer should select the graph and arrangement of information to reduce the number of operations performed. Several researchers have proposed models of mental operations for graph reading—a detailed review of this literature is beyond the scope of this paper [*see (18,21,22,25,64,72–76)*]. For example, Hollands and Spence *(18)* found that trends are best analyzed with line graphs than with a series of pie charts. When estimating trends with line graphs, people can use a slope estimation procedure; with pie charts, they must perform multiple size discriminations between pie slices. However, when estimating proportions, a person performs a ratio estimation task with pie charts, but must perform multiple summation operations before estimating the ratio with line graphs. In addition to preattentive perceptual processes [e.g., detection, discrimination, and identification *(77)*], ratio and summation operations are likely to be common mental operations. In sum, reducing the number of operations will help reduce errors in judgment and in the time needed to perform the task.

Data–Ink Ratio

According to Tufte *(14)*, the amount of ink used to display features other than data should be kept at a minimum. For example, Fig. 10 shows a low data–ink ratio—there is more ink devoted to background and other variables than to the data. In agreement with the recommendations by Tufte *(14)*, Gillan and Richman *(78)* found that participants had faster response times and were more accurate when the data–ink ratio was high than when it was low. In addition, integrated tasks (e.g., global comparisons or synthesis judgments) appear to be more affected by the data–ink ratio than are focused tasks (e.g., selecting the value of a data point). These results suggest that graphs depicting risk data should avoid background or elements that divert attention away from the data (e.g., illustrating the health risks associated with smoking and having a house in the background as shown in Fig. 3). This need to highlight the most relevant data is in keeping with Tufte's other recommendations that are to 1) erase nondata ink, within reason; 2) erase redundant data–ink, within reason; and 3) maximize data density and the size of the data matrix (i.e., how many data points), within reason.

Use of Multiple Graphs

Thus far, our discussion has centered on the construction of single graphs. However, at times, sets of related data must be communicated—when there are several outcomes affected by a single factor (e.g., how smoking affects lung cancer, heart disease, stroke, etc.). In these situations, the graph designer must focus on the relationship between successively viewed graphs, in addition to optimizing the format for each graph. Keep in mind the following three issues:

- *Highlight differences.* Help the observer see the changes or differences from graph to graph in the legends or symbols. For example, if different outcomes (y-axis) are presented as a function of the same factor (x-axis), the designer should highlight the label related to the outcomes.

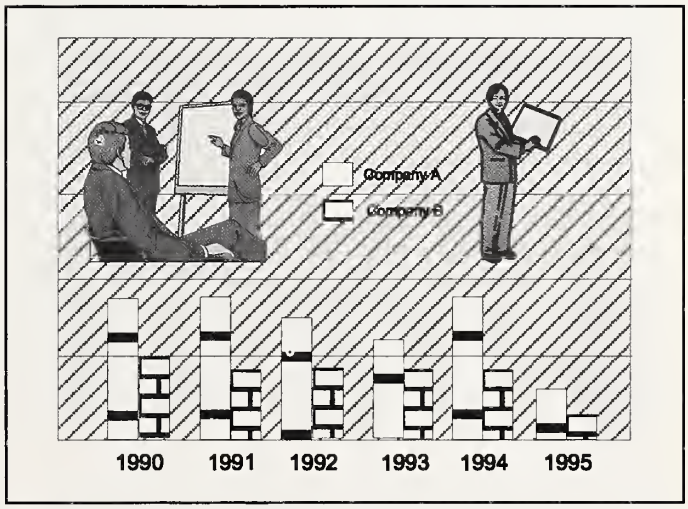


Fig. 10. A graph with a low data–ink ratio. Notice the amount of ink devoted to objects that do not contain the data of interest (pictures, busy background, horizontal grid lines, patterned fills on the bars, etc.).

- *Use distinctive legends.* Legends of similar graphs should highlight their distinctive features.
- *Be consistent.* When plotting the same data in different ways or plotting different data as a function of the same variables, keep elements of the graphs (e.g., colors and legends) consistent. For example, if a variable is highlighted with a particular color in one graph, the same color should be used to depict that variable in another graph.

In addition to the above guidelines, the graph designer should consider at least two additional issues.

Problem of Communicating Small-Probability Events Visually

Although people may have an intuitive understanding of everyday probabilities (e.g., .50), they may fail to comprehend the magnitude of small-probability events (e.g., .0003). One solution is to change probabilities into frequencies (e.g., 3 out of 10 000). Frequency information conforms more readily to people's intuitive assessment of probabilistic occurrences in nature [*see (79)* for review].

Moreover, given that most risk events that threaten our lives have a low probability of occurrence during any particular month or year, it may be more meaningful to communicate aggregate (i.e. cumulative) risk over time *(80)*. For example, whereas the chance of being injured in an automobile accident is roughly about 1 in 10 000 eaeh time a person drives, the chance of being injured at least once during a lifetime of driving is about 1 in 3 *(7)*.

Audience

Knowing the characteristics of the audience is vital to the success of risk communication efforts *(81)*. A critieal question is whether the audience can comprehend the communication [e.g., reading level or preferred method of communication *(81,82)*]. In this respect, two variables may be potentially important: familiarity and experience with graphical displays and how facile people are with mathematical operations and concepts [i.e., numeracy *(11–13)*].

The ability to decipher and comprehend graphs may differ as a function of a person's greater familiarity and working experience with a display. Although performanee with graphs and other visuals should improve with experience and some studies *(19,53)* support this, other studies *(16,65)* show that technical training has no effect on the audience's performance using graphs. Until further research clarifies these relationships further, decisions to use a specific graph to communieate risk should not be based largely on the audience's familiarity with the graph, but more by how well the graph is suited to the task.

In addition, numeracy has been shown to affect risk perceptions. For example, less numerate individuals overestimate personal risks of dying of breast cancer and are less able to accurately gauge the magnitude of risk reduction from mammograms than do more numerate individuals *(11,13)*. Less numerate individuals may stand to benefit more from the visual communication of risk magnitudes and uncertainties than more numerate individuals. However, empirical tests of such effects have yet to be conducted.

Recommendations for Future Research

The visual communication of cancer risk is still in its infancy. In this section, we suggest areas of research that should aid our understanding of these processes.

Conduct Research Linking Graphical Perception to Risk Perception

Research on graphical perception has identified basic graphical forms that should be ideally suited to convey risk information. This research would first identify how basic graphical forms affect risk perceptions along such dimensions as magnitude and accuracy of risk. For example, does varying the size of pie charts and their constituent parts (e.g., number and size of slices) affect perceived risk? This work would capitalize on and integrate findings in the disciplines of psychophysics and graphical perception. Subsequent research could identify how varying graphical elements such as type of numbers (e.g., percents versus frequencies), labels, color, and so forth influences perceived risk.

Determine Effects of Preference for and Experience With Displays on Risk

This information may prove useful when tailoring risk messages to a particular person or audience [*(83); see* also Rimer and Glassman *(84)*, in this monograph]. Ideally, individuals should prefer graphical formats that are consistent with the computational algorithm (e.g., judging magnitude or proportion) suited to a particular graph or visual. People should prefer graphical and visual displays that are consistent with their schemas of what a particular graph or visual is supposed to do.

Establish How Graphs Affect Perceived Severity and Risk

Hazards are often examined with respect to two main components: the nature of the adverse event (i.e., severity, e.g., number of lives lost) and the probabilities that adverse consequences will occur *(85)*. Therefore, it is important to determine how the visual will affect the perceived severity, the probabilities, or both. For example, graphs that appeal to our emotions and arouse thoughts and feelings related to experienced or imagined negative events may initially and most powerfully affect perceived severity. In turn, severity may affect perceived risk by increasing the ease with which negative events are imagined to occur *(86)*.

If the aim of the risk communication is to persuade and initiate behavior change, then it will become important, from a theoretical perspective, to disentangle these effects. For example, several models of health behavior suggest that behavioral change can be predicted by a severity-by-risk interaction, such that hazards that are both severe and highly probable should instigate greater behavior change than hazards that are not very severe and highly probable or very severe and not highly probable [*see (87)* for review].

Establish How Risk Characteristics Interact With Visual Displays to Affect Risk Perceptions

Research using the psychometric paradigm has examined several characteristics of perceived risk, such as whether the risk is voluntary or involuntary, chronic or catastrophic, common or dreaded, injurious or fatal, controllable or not controllable, and old or new [*see (88)* for review]. Perception of risk is greater for those events in which adverse risks are dreaded, uncontrollable, catastrophic, fatal, and that affect future generations. In what ways might these characteristics of risk interact with graphical displays? Perhaps graphs that display such risks might receive more detailed scrutiny (i.e., are more salient), changing the kinds

of mental operations applied to the display. Do displays that communicate dreaded, uncontrollable, catastrophic, or fatal risks engender greater imagery and then affect the risk depicted in a graph?

Determine the Range of Probabilities Conveyed Most Effectively

Are graphical displays most effective for communicating small probabilities, or are they effective at communicating risk at all ranges (i.e., probabilities from 0–1)? For example, the work by Stone et al. *(43)* revealed that visual displays affected how risk-averse participants were via willingness to pay for products using small but not larger probabilities. This study suggests that graphical displays may be powerful aids for conveying the risk of an unlikely event, and people may prefer to have a graphical display in those contexts.

Determine Which Visuals Are Most Effective to Communicate Cancer Risk Uncertainty

We have yet to begin researching how providing visual displays of cancer risk per se affects risk perception, decision-making processes, and, ultimately, behavior. For example, does the addition of a visual display increase or decrease the credibility of the information? Are people less likely to take action when uncertainty information is provided in a graphical format compared with when it is not? Given the various definitions and sources of uncertainty *(89,90)*, this issue deserves more attention.

Use Standard Criteria

To better understand how the addition of graphs improves on other modes of communicating risk (e.g., narratives or numbers only), there is a need for standard evaluation criteria [e.g., those devised by Weinstein and Sandman *(28)*]. Standard criteria would allow conclusions to be drawn across various studies.

Need Multidisciplinary Research

Given the multidimensionality of risk, collaborations between various disciplines and organizations are needed. Working collaboration between experts in human factors, psychology, sociology, psychophysics, graph perception, and the mass media is likely to lead to more integrative and novel approaches than research within a single discipline.

Provide Detailed Accounts of Materials Used to Communicate Risk Visually

In this endeavor, we hope editors will increase the number of graphical and other visuals that can be added to journal articles and book chapters. This increase should enhance tremendously the ability to make comparisons between graphical displays.

Conclusions

The graphical perception literature shows that certain graphs are well suited for specific tasks and provides insight into what to avoid (e.g., use of volume or area and low data–ink ratio). Admittedly, this extant literature has not been applied specifically for communicating cancer risk information and should be explored within this context. Indeed, there is a small but growing amount of experimental literature showing that graphs can affect perceived risk, intentions, and possibly behaviors. For example, the use of the risk ladder shows promise. It is now time for researchers of risk communication to step forward to advance the field of visual risk communication. We hope that this review provides insight into the issues that should be considered when visually displaying risk and spurs researchers to meet the challenges posed by the field of visual risk communication.

References

(1) Covello VT, von Winterfeldt DV, Slovic P. Risk communication: a review of the literature. Risk Abstr 1986;3:171–82.

(2) Fischhoff B. Risk perception and communication unplugged: twenty years of process. Risk Anal 1995;15:137–45.

(3) National Research Council. Improving risk communication. Washington (DC): National Academy Press; 1989.

(4) Slovic P. Informing and educating the public about risk. Risk Anal 1986; 6:403–15.

(5) Halpern DF, Blackman S, Salzman B. Using statistical risk information to assess oral contraceptive behavior. Appl Cogn Psychol 1989;3:251–60.

(6) Siegrist M. Communicating low risk magnitudes: incidence rates expressed as frequency versus rates expressed as probability. Risk Anal 1997;17: 507–10.

(7) Slovic P, Fischhoff B, Lichenstein S. Accident probabilities and seat belt usage: a psychological perspective. Accid Anal Prev 1978;10:281–5.

(8) Stone ER, Yates JF, Parker AM. Risk communication: absolute versus relative expressions of low-probability risks. Organ Behav Hum Decis Process 1994;60:387–408.

(9) Stone ER, Yates JF, Parker AM. Effects of numerical and graphical displays on professed risk-taking behavior. J Exp Psychol Appl 1997;3: 243–56.

(10) Yamagishi K. When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. Appl Cogn Psychol 1997;11: 495–506.

(11) Black WC, Nease RF, Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst 1995;87:720–31.

(12) Paulos JA. Innumeracy: mathematical illiteracy and its consequences. New York (NY): Vintage Books; 1990.

(13) Schwartz L, Wolosin S, Black WC, Welch HG. The role of numeracy in understanding the benefits of screening mammography. Ann Intern Med 1997;127:966–72.

(14) Tufte ER. The visual display of quantitative information. Cheshire (CT): Graphics Press; 1983.

(15) Tufte ER. Envisioning information. Cheshire (CT): Graphics Press; 1990.

(16) Cleveland WS, McGill R. Graphical perception: theory, experimentation, and application to the development of graphic methods. J Am Stat Assoc 1984;70:531–54.

(17) Shamo MK, Meyer J, Gopher D. Predicting values from tables and graphs. Proceedings of the 40th annual meeting of the Human Factors and Ergonomics Society; 1996 Sept 2–6; Philadelphia. Santa Monica (CA): Human Factors and Ergonomics Society; 1996. p. 1151–4.

(18) Hollands JG, Spence I. Judgments of change and proportion in graphical perception. Hum Factors 1992;34:313–34.

(19) Meyer J, Shinar D, Leiser D. Multiple factors that determine performance with tables and graphs. Hum Factors 1997;39:268–86.

(20) Schutz HG. An evaluation of formats for graphic trend displays (experiment II). Hum Factors 1961;3:99–107.

(21) Hollands JG, Spence I. Judging proportion with graphs: the summation model. Appl Cogn Psychol 1998;12:173–90.

(22) Simkin D, Hastie R. An information processing analysis of graph perception. J Am Stat Assoc 1987;82:454–65.

(23) Spence I, Lewandowsky S. Displaying proportions and percentages. Appl Cogn Psychol 1991;5:61–77.

(24) Kosslyn SM. Understanding charts and graphs. Appl Cogn Psychol 1989; 3:185–226.

(25) Pinker S. A theory of graph comprehension. In: Freedle R, editor. Artificial intelligence and the future of testing. Hillsdale (NJ): Erlbaum; 1990. p. 73–126.

(26) Eagly AH, Chaiken S. The psychology of attitudes. Fort Worth (TX): Harcourt Brace College Publishers; 1993.

(27) Nisbett RE, Ross L. Human inference: strategies and shortcomings of social judgment. Englewood Cliffs (NJ): Prentice-Hall; 1980.
(28) Weinstein ND, Sandman PM. Some criteria for evaluating risk messages. Risk Anal 1993;13:103–14.
(29) Henderson M. Living with risk: the choices, the decisions. Chicester (UK): John Wiley and Sons; 1987.
(30) United States Environmental Protection Agency. Indoor air and radiation. In: A citizen's guide to radon. 3rd ed. Washington (DC): Environmental Protection Agency; 1992 Sept; Report No.: EPA 402K92001.
(31) Smith VK, Desvousges WH, Freeman AM. Valuing changes in hazardous waste risks: a contingent valuation analysis. Vol. I: Draft interim report. EPA Cooperative Agreement No. CR-811075: The benefits of hazardous waste management regulations using contingent valuation. Washington (DC): The U.S. Environmental Protection Agency; 1985. Section 8, p. 13–47.
(32) Covello VT. Risk comparisons and risk communication: issues and problems in comparing health and environmental risks. In: Kasperson RE, Stallen PJM, editors. Communicating risks to the public: international perspectives. Dordrecht (The Netherlands): Kluwer Academic Publishers; 1991. p. 79–124.
(33) Weinstein ND, Sandman PM, Roberts NE. Communicating effectively about risk magnitudes, Phase 1. New Brunswick (NJ): Environmental Communication Research Program, Cook College, Rutgers University; 1989.
(34) Weinstein ND, Sandman PM, Miller P. Communicating effectively about risk magnitudes, Phase Two. New Brunswick (NJ): Environmental Communication Research Program, Cook College, Rutgers University; 1991.
(35) Smith VK, Desvousges WH, Johnson FR, Fisher, A. Can public information affect risk perception? J Policy Anal Manage 1990;9:41–59.
(36) Sandman PM, Weinstein ND, Miller P. High risk or low: how location on a "risk ladder" affects perceived risk. Risk Anal 1994;14:35–45.
(37) Smith KV, Desvouges W, Payne JW. Do risk information programs promote mitigation behavior? J Risk Uncertainty 1995;10:203–21.
(38) Fisher A, McClelland GH, Schulze WD. Communicating risk under Title III of the SARA; strategies for explaining very small risks in a community context. J Air Pollut Control Assoc 1989;39:271–6.
(39) Sandman PM, Weinstein ND, Hallman WK. Communications to reduce risk underestimation and overestimation. Risk Decis Policy 1998;3:93–108.
(40) Sandman PM, Weinstein ND. Communicating effeetively about risk magnitudes: bottom line conclusions and recommendations for practitioners. Washington (DC): Environmental Protection Agency; 1994 Aug; Report No.: 230.
(41) Desvouges SH, Smith VK. Focus groups and risk communication: the 'science' of listening to data. Risk Anal 1988;8:479–84.
(42) Johnson BB, Slovic P. Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. Risk Anal 1995; 15:485–94.
(43) Stone ER, Rush CJ. Risk communication: the effectiveness of graphical modes depends on the risk magnitude. Poster presented at the annual meeting for the Society for Judgement and Decision-making; 1997 Nov; Philadelphia (PA).
(44) Stephenson MT, Witte K. Fear, threat, and perceptions of efficacy from frightening skin cancer messages. Public Health Rev 1998;26:147–74.
(45) Mazur DJ, Hickman DH. Patients' and physicians' interpretations of graphic data displays. Med Decis Making 1993;13:59–63.
(46) Kaplan RM, Hammel B, Schimmel LE. Patient information processing and decision to accept treatment. J Soc Behav Pers 1985;1:113–20.
(47) Weinstein ND, Sandman PM, Hallman WH. Testing a visual display to explain small probabilities. Risk Anal 1994;14:895–7.
(48) Baty BJ, Venne VL, McDonald J, Croyle RT, Halls C, Nash JE, et al. BRCA1 testing: genetic counseling protocol development and counseling issues. J Genet Couns 1997;6:223–44.
(49) National Cancer Institute, Office of Cancer Communications. How the public perceives, processes, and interprets risk information: findings from focus group research with the general public. Bethesda (MD): National Cancer Institute; 1998 June; Report No.: POS-T086.
(50) Hampson SE, Andrews JA, Lee ME, Foster LS, Glasgow RE, Lichtenstein E. Lay understanding of synergistic risk: the case of radon and cigarette smoking. Risk Anal 1998;18:343–50.
(51) Lathrop RG. Perceived variability. J Exp Psychol 1967;73:498–502.
(52) Legge GE, Gu YC, Luebker A. Efficiency of graphical perception. Percept Psychophys 1989;46:365–74.
(53) Ibrekk H, Morgan GM. Graphical communication of uncertain quantities to nontechnical people. Risk Anal 1987;7:519–29.
(54) Keeney RL, von Winterfeldt D. Improving risk communication. Risk Anal 1986;6:417–24.
(55) Fiskel J. The impact of artificial intelligence on the risk analysis profession. Risk Anal 1987;7:277–80.
(56) Hollands JG, Dyre BP. Bias in proportion judgments: the cyclical power model. Psychol Rev. In press 1999.
(57) Gillan DJ, Wickens CD, Hollands JG, Carswell CM. Guidelines for presenting quantitative data in HFES Publications. Hum Factors 1998;40: 28–41.
(58) Poulton EC. Geometric illusions in reading graphs. Percept Psychophys 1985;37:543–8.
(59) Tversky B. Distortion in memory for visual displays. In: Ellis SR, Kaiser MK, Grunwald AJ, editors. Pictorial communication in virtual and real environments. London (UK): Taylor and Francis; 1991. p 61–75.
(60) Stevens SS, Galanter EH. Ratio scales and category scales for a dozen perceptual continua. J Exp Psychol 1957;54:377–411.
(61) Stevens SS. Psychophysics: introduction to its perceptual, neural, and social prospects. New York (NY): John Wiley and Sons; 1975.
(62) Varey CA, Mellers BA, Birnbaum MH. Judgments of proportions. J Exp Psychol Hum Percept Perform 1990;16:613–25.
(63) Sparrow JA. Graphical displays in information systems: some data properties influencing the effectiveness of alternative forms. Behav Inf Technol 1989;8:43–56.
(64) Carswell CM. Choosing specifiers: an evaluation of the basic tasks model of graphical perception. Human Factors 1992;34:535–54.
(65) Cleveland WS, McGill R. An experiment in graphical perception. Int J Man-Mach Stud 1986;25:491–500.
(66) Wainer H. Visual revelations: graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot. New York (NY): Copernicus; 1997. p. 104–10.
(67) Wickens CD, Carswell CM. The proximity compatibility principle: its psychological foundation and relevance to display design. Hum Factors 1995;37:473–94.
(68) Chernoff H. The use of faces to represent point in K-dimensional space graphically. J Am Stat Assoc 1973;68:361–8.
(69) Jacob R, Egreth HE, Bevan W. The face as a display. Hum Factors 1976; 18:189–200.
(70) MacGregor D, Slovic P. Graphical representation of judgmental information. Hum-Comput Interact 1986;2:179–200.
(71) Carswell CM. Graphical information processing: the effects of proximity compatibility. Proceedings of the Human Factors and Ergonomics Society 34th annual meeting; 1990 Oct 8–12; Orlando. Santa Monica (CA): Human Factors and Ergonomics Society; 1990. p. 1494–8.
(72) Casner SM. A task-analytic approach to the automated design of graphic presentations. ACM Trans Graphics 1991;10:111–51.
(73) Gillan DJ. Visual arithmetic, computational graphics, and spatial metaphor. Hum Factors 1995;37:766–80.
(74) Gillan DJ, Lewis R. A componential model of human interaction with graphs: 1. Linear regression modeling. Hum Factors 1994;36:419–40.
(75) Gillan DJ, Lewis R. A componential model of human interaction with graphs II. Effects of the distances among graphical elements. Proceedings of the Human Factors and Ergonomics Society 36th annual meeting; 1992 Oct 12–16; Atlanta. Santa Monica (CA): Human Factors and Ergonomics Society; 1992. p. 365–8.
(76) Lohse GL. A cognitive model for understanding graphical perception. Hum-Comput Interact 1993;8:353–88.
(77) Winn W. Contributions of perceptual and cognitive processes to the comprehension of graphics. In: Schontz W, Kulhavy RW, editors. Comprehension of graphics. Amsterdam (The Netherlands): North-Holland; 1994. p. 3–27.
(78) Gillan DJ, Richman EH. Minimalism and the syntax of graphs. Hum Factors 1994;36:619–44.
(79) Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: frequency formats. Psychol Rev 1995;102:684–704.
(80) Doyle JK. Judging cumulative risk. J Appl Soc Psychol 1997;27:500–24.
(81) Lundgren RE. Risk communication: a handbook for communicating

environmental, safety, and health risks. Columbus (OH): Batelle Press; 1994.
(82) Health literacy: report of the Council on Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA 1999;281:552–7.
(83) Rimer BK, Glassman B. Tailored communications for primary care settings. Methods Inf Med 1998;37:171–7.
(84) Rimer BK, Glassman B. Is there a use for tailored print communications in cancer risk communication? J Natl Cancer Inst Monogr 1999;25:000–00.
(85) Hannson SO. Dimensions of risk. Risk Anal 1989;9:107–12.
(86) Tversky A, Kahneman D. Availability: a heuristic for judging frequency and probability. Cogn Psychol 1973, 5:207–32.
(87) Weinstein ND. Testing four competing theories of health-protective behavior. Health Psychol 1993;12:324–33.
(88) Slovic P. Perceptions of risk: reflections on the psychometric paradigm. In: Krimsky S, Golding D, editors. Social theories of risk. Westport (CT): Preager; 1992. p. 181–216.
(89) Rowe WD. Understanding uncertainty. Risk Anal 1994;14:743–50.
(90) Babrow AS, Kasch CR, Ford LA. The many meanings of uncertainty in illness: toward a systemic accounting. Health Commun 1998;10:1–23.
(91) Messaris P. Visual persuasion: the role of images in advertising. Thousand Oaks (CA): Sage Publications; 1997.
(92) Apaiwongse TS. Facial display of environmental policy uncertainty. J Bus Psychol 1995;10:65–74.
(93) Denes-Raj V, Epstein S. Conflict between intuitive and rational processing: when people behave against their better judgement. J Pers Soc Psychol 1994;66:819–29.
(94) Denes-Raj V, Epstein S, Cole J. The generality of the ratio-bias phenomenon. Pers Soc Psychol Bull 1995;21:1083–92.
(95) Kirkpatrick LA, Epstein S. Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. J Pers Soc Psychol 1992;4:534–44.

Notes

[1]For a discussion as to how visuals other than graphics (e.g., pictures) might be used to persuasively communicate disease risk, the reader is referred to the text titled *Visual Persuasion: The Role of Images in Advertising (91).* For example, pictures may be effective, because they realistically capture, and hence provide, convincing data pertaining to real-life examples related to risk (e.g., disease severity).

[2]Smith et al. *(35,37)* tested how six formats to communicate radon risk affected risk perceptions and mitigation intentions. Of import, they varied in two formats the advice (i.e., tone) given in terms of how homeowners should interpret their actual radon levels. In the Command version, homeowners were told to follow the Environmental Protection Agency (EPA) guidelines for levels warranting action. For the Cajole version, participants were encouraged to use their own judgments and evaluations and used several standards (e.g., National Council on Radiation Protection), including the EPA's, to evaluate their radon levels. Similar to the results of Weinstein and colleagues *(33),* their outcomes suggest that providing directive advice encourages mitigation.

[3]Although not applied to risk communication, some researchers have made use of Chernoff face displays for portraying complex multivariate data *(68–70)* and as a method of conveying attitudinal uncertainty toward environmental policies *(92).* In these displays, facial features (eyes, nose, and mouth) vary with the magnitude of the variable they display.

[4]In decision-making trials, some researchers have used different colored jelly beans, rather than dots, to represent probabilities. One common observation is the ratio–bias phenomenon. People judge the occurrence of a low-probability event as less likely when the same probabilities are presented by the ratio of smaller (e.g., 1 in 20) than larger (e.g., 10 in 200) numbers *(93–95).*

[5]The National Cancer Institute *(49)* also used focus groups to test others visuals depicting the relative risks linking lung cancer with smoking. These visuals used blackened lungs and hospital beds to illustrate that smokers are significantly more likely (i.e., 10 times more likely) to get lung cancer than nonsmokers. Although most participants understood that smoking was linked with lung cancer, several interpretative problems emerged. The reader is referred to the publication by the National Cancer Institute *(49)* for the detailed results.

Supported by Public Health Service grant 5U19CA72099-03 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

We thank Eric Stone, Paul Slovic, Frank Yates, Caryn Lerman, Kerry Smith, William Desvousges, Melody Carswell, Joachim Meyer, and Neil Weinstein for providing suggestions, resources, and references. We thank Jimmy Rosen for his help in creating some of the graphics. We thank Barbara Rimer for her comments on an earlier version of this manuscript.

Risky Business—Communicating Scientific Findings to the Public

Jeffrey Levine

Communicating the risk of anything can be hazardous to the truth, as well as to the goal of good understanding.

When a journalist and a scientist agree to discuss a research finding, the conversation usually involves some description of risk—the probability that a disease will or will not occur, based on a specific set of assumptions. For example, what is the likelihood that a breast cancer gene will cause a tumor? What are the chances a diet high in cholesterol will lead to cardiovascular problems?

These highly technical questions hinge on a number of fine points that add up to credible information in the researcher's mind and, hopefully, a story for the reporter. However, a built-in chasm exists between the scientific worldview and the journalistic approach to handling information.

Risk to the scientist is ultimately a tool to measure and compare data, a statement of probability. For journalists, risk is a measurement of news value. The greater the risk of anything, the bigger the story. Is there a way of reconciling these disparate approaches?

Journalists and scientists must stop viewing each other as adversaries, except in extreme circumstances such as scientific fraud. News people are not health educators, but their primary goal should be to communicate meaningful information fairly and truthfully in a way that will allow the public to make appropriate decisions about their health. If a few papers are sold along the way, so be it. Everybody benefits.

Unfortunately, reporters and their editors are often at scal-

Correspondence to: Jeffrey Levine, M.J., Washington DC Bureau, Medcast Networks, National Press Building, Suite 1290, 529 14th St., N.W., Washington, DC 20045 (e-mail: jlevine@gnncast.net).

© Oxford University Press

pels' points with scientists whose findings they are describing. Why? What is the issue? Scientists and reporters, who represent the public, have drastically differing views of risk.

Most would consider an expedition into the wilderness a risky prospect. But for Meriwether Lewis, the risks of civilization proved greater than those of the unknown. He survived the tests of the western United States for 3 years but committed suicide shortly after his return *(1)*.

The British explorer George Mallory chose to undertake what proved for him to be the ultimately fatal risk of Mt. Everest simply, as he told a *New York Times (2)* reporter in 1923, "Because it's there." Risk-taking in the lofty contexts of great achievements is an essential virtue.

However, in a more mundane way, risk may simply mean the chance of getting caught, say, when one's car is in the cross hairs of the policeman's radar gun or when a politician's personal peccadilloes are uncovered by the news media. If risk is simply going to Las Vegas for some, or cracking the genome's mysteries for others, then explaining research findings meaningfully to both extremes is challenging, if not impossible.

Another issue is that an intensely politieal atmosphere often surrounds the reporting of disease and treatments. If a patient with human immunodeficiency virus wants a new treatment, the notion of risk shifts significantly, based on the certainty of death associated with the infection. Desperate problems require desperate measures that move the yardsticks of risk to their outer boundaries.

Statistics convince scientists of risk, but they do not sell papers or get people to watch television newscasts. A good journalist knows that he or she has to put the relevant facts in the story, but the "spoonful" of sugar is the anecdote. Variously known as the sob story, or the emotional "spike," the anecdote is an indispensable element in a piece about medicine or science, which otherwise would be impenetrable to the average person.

The anecdote is the way into the story for most reporters: Get the audience to commit to the person who has the problem, then explain it. Of course, there is a danger here—the reader or viewer may not get past the personal "tale of woe" to consume the nugget of information. In addition, because an anecdote is just a single case, it may or may not fairly represent the facts. Still, this is the reality of science reporting today, particularly on television.

Rarely, if ever, will the reporter simply lay out the scientific findings or explain *P* values, odds ratios, or relative risks. Techno-babble is out. Touchy-feely is in. Dumbing down has reached the point that, as a network medical correspondent was told, *placebo* was too difficult a word to use in a story, or that even the word "risk" itself should be expunged from copy as being difficult to comprehend.

Well, the reporter could understand saying "dummy drug" instead of placebo. However, eliminating risk from medical coverage could reduce the stories to the intellectual equivalent of absolute zero.

Yet medical and science reporters are constantly faced with the challenge of explaining their topics without explaining them away. That is why the notion of risk is often so overblown—it is a device used to grab a viewer or reader. Indeed, risk is the kernel of the story, but it is often expressed without the nuance or science that earned the study favorable peer review and acceptance by a journal.

Over the years, certain stories stand out as examples of, as it were, the risks inherent in reporting on risk. Although they do not necessarily resolve the dilemma, the issues are interesting anecdotes on how the question of risk plays out in the real world of news.

The safety of silicone gel-filled breast implants occupied much of this reporter's time in the early 1990s. The controversial devices had been on the market for some 30 years by the time the U.S. Food and Drug Administration (FDA) gained regulatory authority over them. When ordered by then FDA Commissioner David Kessler to prove that the devices were safe in widely covered public hearings, the industry was unable to satisfy the critics, many of whom were litigating and winning multimillion dollar damage suits.

What made the story particularly poignant is that women were both the victims and the benefactors of the implants. Many survivors of breast cancer hailed the devices for restoring their bodies and their self-esteem. Other embittered recipients believed the implants, hailed by physicians as low-risk devices, had wrecked their health. Large amounts of airtime were consumed with dramatic footage about a story that combined sex, science, and the alleged abuses of corporate power in unusual and compelling ways.

It is not as though journalists did not feel justified in reporting what they knew about the implants. "By now, with 20 years of experience in risk reporting, reporters approach the subject with greater independence, and indeed, a certain cynicism about corporate behavior," writes Professor Dorothy Nelkin *(3)* of New York University in an essay titled "Reporting Risk: The Case of Silicone Breast Implants."

Nelkin, a specialist in sociology and law, also says in the piece, "The silicone implant fiasco followed on the heels of a series of cases in which companies have been found to obfuscate known risks—for example, the risks of asbestos, of dioxin, and of cigarette smoking."

The fact that the multibillion dollar implant settlement is still being debated some 6 years later is a testimonial to the difficulty in reaching a scientific and legal consensus about the product. It also suggests that the news media really did not do as well as they should have with the story, but then getting good science took a long time. Stories are no better than the research from which they are derived.

If the implant coverage was an emotional morass, the reporting of the tobacco debate offered more than a modicum of light to accompany its intense heat. David Kessler again played a key role. As discussed in many interviews with this reporter, beginning in 1994, Kessler started his move to regulate tobacco and cigarettes as a combination drug and drug-delivery device.

For reporters, this move was a "slam dunk." What is the risk of smoking? Kessler's mantra was that 3000 children take up the habit every day, and eventually1000 of them will die of smoking. Abetted by a multibillion dollar ad effort, the tobacco industry was accused by Kessler and public health officials like former Surgeon General C. Everett Koop, M.D., of attempting to seduce children with an array of cartoon charaeters and merchandising promotions aimed directly at an underage audience *(4)*. This reporter heard such arguments made directly on countless occasions.

The antitobacco juggernaut was also joined by advocacy groups, like the American Medical Association and the Ameri-

can Cancer Society. Whereas the breast implant manufacturers were seen as bumbling or, in many cases, clumsily covering up questions of safety about the product, the tobacco industry was portrayed as a sinister and conspiratorial enterprise that knew the evils of its efforts and denied them totally. Nowhere was this view more reinforced than in congressional hearings in 1994, witnessed by this reporter, in which tobacco executives swore under oath that their products were not dangerous nor addictive. Such obvious hubris was greeted with a widespread gasp of incredulity both by journalists and the public.

Only part of this view can be attributed to the accurate reporting of scientific and medical risk. But it could be argued that, given better data to work with and a more engaging target, the media were able to do a better job of communicating a real, immediate risk as well as an abiding hazard.

Even though tobacco regulation awaits final judgment in the Supreme Court, and the tobacco industry has settled for a smaller version of an earlier deal with state attorneys general, the public's heightened perception of risk and blame drove a once untouchable industry into a previously unthinkable compromise.

Things would be simpler, albeit more dull, if scientific unanimity was available on a given topic. Of course, that is not the case anymore than Arabs and Israelis are likely to find common ground on the major issues dividing them.

For instance, consider the debate over mammography guidelines. The issue hinged on real scientific questions about whether screening of women in their 40s actually prevented a significant number of breast tumors or simply exposed women to medical procedures that might prove to be harmful in their own right. However, as with the implant debate, the issue was every bit as emotional as empirical.

When a National Institutes of Health (NIH) panel *(5)* said in January 1997 that it could not find sufficient evidence to recommend mammography for women in their 40s, it created a fire storm. A resolution that passed in the U.S. Senate attacked the conclusion and urged the National Cancer Institute (NCI) to reconsider the actions. Facing possible budgetary reprisals from Congress, the NCI took another look at the evidence.

Within months, a new and broader screening guideline was made from another panel that looked at the risks and benefits of mammography in another light. Reporters were not surprised, but many were impressed by NCI Director Richard Klausner's ability to navigate the storm that at its core was not simply a matter of piling up the numbers to support one case or another.

The affair is well documented by Steven Woolf and Robert Lawrence of the Medical College of Virginia and The Johns Hopkins School of Hygiene and Public Health, respectively, in a *Washington Post* Op-Ed piece in May 1997 *(5)*. But the controversy stands out in the mind of this reporter who witnessed the government's tortured efforts to reach consensus.

Recalling the hothouse environment of both the initial meeting and then the review, clearly sentiment as much as science was driving the debate. Back at the newsroom, intense interest, if not outright glee, was growing that experts, interest groups, and legislators were engaged in guerilla warfare.

Many stories from that era, certainly this reporter's included, inevitably contained a certain bias toward early mammography. That bias was not intentional, but the prevention case seemed to make a better journalistic argument. Even if a risk exists in doing something, the psychology dictates that it is better to take that risk than do nothing to prevent cancer. This argument is not based on science but on the very real human need to take action in the face of a terrible threat.

Reporters report facts and arguments, but some facts have greater weight than others. Even reporters who tried to do a better job were no doubt pressured by editors who wanted to "shape" the story to satisfy their own ends, namely, wider circulation or viewership.

Here is how this reporter presented the story at its high point in March 1997 in an excerpted script prepared for national broadcast *(6)*.

> (Narration) The National Cancer Institute was under intense political and medical pressure to change its view that early mammography wasn't necessary.
>
> (Sound bite with Richard Klausner, M.D., NCI Director) Screening mammography is capable of saving lives, of reducing mortality from breast cancer. The question that has been difficult to resolve is when women should begin.
>
> (Narration) Earlier the American Cancer Society recommended starting annual mammography at age 40.

Clearly, the pressure for the new guideline had been building and was ultimately irresistible. In an earlier script on the same subject, I included the following comment from Director Klausner: "As I said at the end of the consensus conference, we can't create certainty where there is no certainty." However, that was clearly the problem—certainty was demanded by the public and the news media, and science was not cooperating.

At some point, we were repeatedly told that a better technology will be available for detecting breast tumors that will render the debate moot. Until then, reporters and their critics must agree that some stories are ultimately explained in a bigger context no matter what the science says.

When questions of public policy dominate the debate, the communication of real risks and benefits are often a secondary consideration. In these conflicts, emotion will carry the day more often than not in spite of a reporter's efforts to bring reason and logic to bear.

In the heated atmosphere of reporting risk, are there steps reporters and scientists can take to limit confusion and enhance communication? Indeed, there are, and they are often practiced, though not to the extent they should be. Consider the following steps:

1) Scientists bear much of the burden. They need to realize when they are dealing with journalists that they are not talking to their peers. Clarity and directness should be the goals of such discussions. Sometimes, researchers should rehearse their presentations in front of communications professionals if necessary to make sure their message is going to get through.

2) A caveat here—scientists should by no means "spin" their interviews. Nor should they polish themselves to a state of talk-show glibness. But they need to know how to explain, within limits, what are the risks and benefits of a given problem in plain English.

3) Researchers must learn how to demystify media and to avoid viewing an encounter with a journalist as a toxic experience. Some of the better scientists have learned the obvious political lesson that a solid media presentation not only encourages understanding but also raises the individual's stature in the community. This is not ego. It is simply smart business, particularly in a media-saturated environment. In other words, it is better to be a confident advocate than a hapless apologizer.

4) The journalist's posture must be more restrained and cautious—difficult qualities in an environment soaked with sensation of every sort. Yet, reporters and editors must look thoughtfully at every new study and ask themselves if it merits raising the opposing semaphores of alarm or of congratulation. Angiogenisis is an exciting approach to cancer but then so was interferon 15 years ago.

5) Whether reporters understand the science of statistics or not, sooner or later those who work the health and medical beats can tell when a study rises above the horizon line. Acting on such notions often requires courage; but if no distinctions are drawn, little if any value will be reported.

6) Journalism, like medicine, is becoming more specialized. Reporters who cover health and science as a beat are more inclined to get it right. They know the history of an issue and have a better chance of keeping one story, no matter how sensational, in perspective. Scientists should make an effort to work with reporters who have some appreciation for the material.

Meanwhile, viewers and readers are becoming ever more skeptical. Does everything cause cancer? A1998 report that declared alcohol a human carcinogen is unlikely to affect sales or consumption of the product, at least not much. If by sensationalizing stories journalists are "crying wolf," ultimately no one is going to pay attention.

Sorting out the real from the hypothesized risks is becoming a greater challenge. A cartoon *(7)* in a recent *New Yorker* shows a dog on a couch talking to his canine therapist. The caption is, "I do what they tell me, I eat what they give me. How do I know they're not a cult?"

Even doing all the right things is not enough if confidence in the integrity of the information is consistently undermined.

A plane crash will certainly generate more attention than the fact the meals eaten by the passengers generally constitute a longer-term threat to health than the transitory risks of flying.

In an essay titled "Reporting on Risk: How the Media Portray Accidents, Diseases, Disasters, and Other Hazards," Eleanor Singer and Phyllis Endreny *(8)* observe that the media tend to focus on "catastrophic" accidents. "Such accidents rank near or at the top in terms of media attention, even though with one exception (automobile accidents) they do not result in a large number of deaths per year. Thus, media definitions of risk are based on the drama of the single hazardous event," they argue.

So the reporting of risk is a hazardous affair in which dramatics are always endangering relevance and one in which journalists and researchers often find themselves at opposing ends of the information spectrum. Still, like occasional antagonists in a marriage of long standing, the two parties need each other. They will never fully reconcile their differenees, but the benefits of their staying together far outweigh the risks of their falling out.

References

(1) Ambrose S. Undaunted courage. New York (NY): Simon & Schuster; 1996.

(2) Platt S, editor. Respectfully quoted: a dictionary of quotations requested from the Congressional Research Service. Washington (DC): Library of Congress; 1989. p. 327.

(3) Nelkin D. Report risks: the case of silicone breast implants. In: Risk: health, safety, & environment. Vol. 5. Concord (NH): Franklin Pierce Law Center, Risk Assessment & Policy Association; 1994. p. 233.

(4) Orrick DJ. Tackling tobacco. Tobacco and advertising: part I [on line] [cited 1998 Dec 10–11] Available from: http://www.drkoop.com/wellness/tobacco/articles/advertising1.asp

(5) Woolf S, Lawrence R. When politicians play doctor; the mammogram debate shows why it's dangerous. The Washington Post 1997 May 4; C1.

(6) Mammography [news script]. CNN prime news; 1997 Mar 27.

(7) Barsotti C, artist. Dog talking with therapist [cartoon]. New Yorker 1998 Nov 16; 100.

(8) Singer E, Endreny P. Reporting on risk: how the media portray accidents, diseases, disasters, and other hazards. In: Risk: health, safety, & environment. Vol. 5. Concord (NH): Franklin Pierce Law Center, Risk Assessment & Policy Association; 1994. p. 261.

Living Can Be Hazardous to Your Health: How the News Media Cover Cancer Risks

Cristine Russell

For more than two decades, the news media has bombarded the public with often conflicting information about health risks, contributing to an atmosphere of hype and hysteria about cancer and other diseases. Improvements in media reporting of health risks require greater efforts by both those who cover the news and those who create it. Guidelines for bringing more perspective and balance to media coverage of risk are provided. These include putting cancer in context with other diseases, explaining absolute and relative risks, differentiating between individual and population risks, stressing the degree of uncertainty of new research and how it fits with previous data, covering the process as well as end results of science, understanding different media constraints and needs, and taking into account the diverse backgrounds and needs of the target audience—the general public. [Monogr Natl Cancer Inst 1999;25:167–70]

There is little, if any, humor to be found in cancer, the nation's second leading killer, or in the known and suspected causes of this diverse group of diseases. It is a deadly serious subject. But there is some comic relief in cartoons that manage to poke fun and provide insight into difficult problems. Such is the case with a cartoon bearing a facetious headline that rings true: "Today's Random Medical News from the New England Journal of Panic-Inducing Gobbledygook." It features a newscaster announcing the latest findings from "a report released today" involving yet another health scare. His backdrop is a set of three roulette wheels listing seemingly random health risks—from smoking to stress—that may or may not cause a variety of problems—from depression to sexual dysfunction—in everything from rats to men 25–40 years old *(1)*.

Unfortunately, for nearly three decades, media coverage of health risks, particularly cancer risks, has increasingly seemed to follow a wheel-of-fortune approach. Be it hourly, daily, weekly, or monthly, journalists following the latest study spin out a new health risk that often contradicts a previous study and helps contribute to a general feeling of confusion and concern in the general public. The lasting impression, however mistaken, is that virtually every aspect of daily life can indeed be hazardous to your health. The Surgeon General's warning about smoking has become ubiquitous, extending to a seemingly endless list of major and minor hazards with varying degrees of proof about their relevance to human health.

The bombardment of bad news, particularly about purported cancer risks, runs the gamut. The pesticide Alar® and apples. Estrogen replacement therapy and possible breast cancer risk. Benzene contamination of Perrier sparkling water. Second-hand smoke. Asbestos in schools. The artificial sweetener saccharin and cancer in rats. Radon and lung cancer. The list goes on and on.

The coverage of known and suspected cancer risks is not merely a creation of the media, however. There is plenty of blame to go around. The often misleading risk coverage involves a variety of players, including bench scientists, clinicians, universities, manufacturers, public relation firms, medical journals, advocacy groups, lawyers, and politicians.

To understand the current coverage of cancer, it is important to look back to 1971, when the national "war on cancer" was launched. This was the first story I covered as a journalist, and I remember the message sent from official Washington, from the President on down: if we spend enough money on a crash research program, we can win a war against the feared enemy cancer. Some argued that if we can put a man on the moon, surely we can cure cancer. Fortunately, reason prevailed over rhetoric, and cancer research, with greatly increased funding, was kept within the umbrella of the National Institutes of Health *(2)*. Unfortunately, both the media and the medical community have continued to use military metaphors in the coverage of cancer and its risks. It has always been somewhat misleading because the emphasis on who's winning and who's losing creates a body-count approach to cancer in the public eye. The "war" on cancer is, of course, still going on, and while some battles have been won, the casualties remain very high.

In terms of language, the ongoing use of the word "cancer" in a singular manner also continues to give the misleading impression of one disease rather than emphasizing the complexity of more than 100 different diseases with a multitude of different risk factors.

Cancer coverage in the 1970s set the tone for coverage to come. As money was poured into animal testing, reporters faced the challenging new "carcinogen-of-the-week syndrome" and found themselves covering mice more than men. A Robert Mankoff cartoon, showing two mice talking in a cage, captured the moment: "My main fear used to be cats—now its carcinogens" *(3)*. The federal chemical testing program led to a deluge of fragmented, scary, difficult-to-interpret stories that contributed to the feeling among the public that everything causes cancer. New testing methods made it possible to detect ever smaller amounts of carcinogens in the environment. A widely publicized pronouncement by prominent researchers that more than 90% of cancer is "environmental" fueled the fire. In this case, the word "environmental" was misleading because the all-encompassing phrase technically meant everything that was not "genetic," including personal actions like eating and smoking. But because of the emerging environmental movement, the pronouncement was viewed by some as a confirmation that chemical pollutants in the air, land, and water were major cancer culprits. Yet there was no consensus in the scientific community about what the newly documented risks really meant in terms of human health hazards.

Affiliation of author: Freelance writer and special health correspondent, *The Washington Post.*

Correspondence to: Cristine Russell, 43 Huckleberry Ln., Darien, CT 06820

© Oxford University Press

The overemphasis on the environmental causes of cancer was countered to some degree in the late 1970s and in the 1980s by a growing recognition that "lifestyle" was in fact the major cause of most cancers. Campaigns to influence individual behavior and prevent cancer and other diseases gained more attention and "prevention" became a buzzword. Cigarette smoking—long recognized by experts as the leading preventable cause of death and disease in the country—began to receive more attention, aided by government officials like outspoken Health, Education, and Welfare Secretary Joseph A. Califano, Jr., in the Carter administration and charismatic Surgeon General C. Everett Koop in the Reagan administration. Finally, both the health and journalistic communities started to provide more perspective on the behavioral side of cancer risk. The comment by cartoon character Pogo, "We have met the enemy and he is us," hit home *(4)*.

In the 1990s, cancer continued to receive considerable media attention, particularly specific cancers of the breast and prostate. New genetic studies provided insight into which cancers were hereditary and generated potential tools for detecting cancer genes. Changes were also made in the way scientists worked with the media. When I began covering cancer in the 1970s, researchers were reluctant to speak to us and sought to hide behind the shield of press releases and journals. Although that is still true for some, more researchers see the value of communicating with the public through the media. Some doctors have even hired large public relations firms to promote their work, and journalists now find themselves contending with some overenthusiastic medical promoters. As a *New Yorker* cartoon, showing a doctor examining a patient (Fig. 1), put it: "Mr. Wilkins, I believe your condition is going to get us both into the *Journal of the American Medical Association*" *(5)*.

A greater eagerness for publicity is also seen in the promotional efforts of medical journals, many of which now provide advance press release packages and have release dates and times carefully chosen to increase the chance of making the evening news. Institutions, from universities to companies with new medical products to sell, also vie for more media attention. In addition, the legal system, as we have seen with the heated controversy over the safety of breast implants, has increasingly intruded into medical coverage and further polarized already controversial scientific issues.

"Mr. Wilkins, I believe that your condition is going to get us both into the Journal of the American Medical Association.' "

Fig. 1. © The New Yorker Collection 1995 W. B. Park from cartoonbank.com. All rights reserved. Reprinted with permission by *The New Yorker.*

The intense coverage, on a story-by-story basis, has emphasized disagreement over agreement. As one story contradicts another, or experts take extreme sides, the public confidence in both the media and the scientific community diminishes. One of the most famous *New Yorker* cartoons (Fig. 2) about risk shows a tanker truck barreling down a highway, with this message painted on the truck's side: "The scientific community is divided. Some say this stuff is dangerous, some say it isn't" *(6)*. In addition to conflict, there is also uncertainty, particularly in the risk arena. This uncertainty might be called the National Academy of Sciences syndrome, because so many reports on important medical and scientific issues end with the caveat that "more research is needed." One cartoon by Sidney Harris, who is a master at capturing scientific humor, shows two lab-coated researchers in intense conversation: "Granted, we have to do the research, and we can do some research on the research, but I don't think we should be involved in research on research on research" *(7)*.

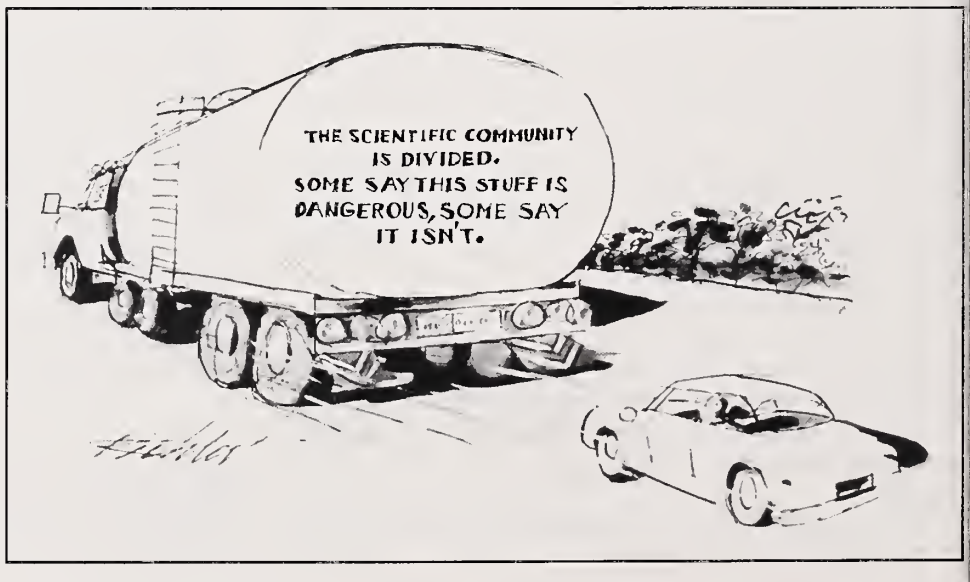


Fig. 2. © The New Yorker Collection 1988 Mischa Richter from cartoonbank.com. All rights reserved. Reprinted with permission by *The New Yorker.*

The problem is that most times the news media, both print and broadcasting, simply cannot wait for the scientists to complete all of the studies needed to reach complete consensus. Instead, we need to give the best information available to the public at a given point in time. But how well do we do that? Certainly, as journalists, we have contributed to the public hysteria over cancer risks. As a group, we do indeed tend to emphasize the dramatic over the mundane, new risks over old ones, and conflict and drama, particularly in the environmental, political, or legal arenas. We seldom report on the "negative" stories about things that do not cause cancer.

Nonetheless, I believe that reporters with better training and experience in medical and science reporting, working with public information offices and scientists that are truly willing to work with the media, ean do a much better job of putting risk coverage into perspective. Instead of reporting each story in a vacuum, many of us are trying harder to present stories involving risk in a framework that emphasizes what we know and what we need to know, who is at risk and who is not at risk, what can be done now in terms of prevention and what needs to be done in the future.

Working to improve media coverage of cancer, whether basic research or epidemiologic studies involving risk, prevention, or treatment, the key element is balance. The following are ten tips that reporters, as well as those who help create the news, should keep in mind to improve coverage.

1) *Put cancer in context.* Although cancer is the nation's second leading killer, it is obviously not the only set of diseases that threatens the public. Heart disease kills more Americans. Women should not be made so fearful of breast cancer that they fail to worry about actions they can take to prevent heart disease. But this should not be a competition. The key is to put cancer in context with other causes of morbidity and mortality.

2) *Stop the "yo-yo" approach to cancer coverage.* Too often stories hype alarming new risks and overpromote promising new findings. Veteran *Washington Post* science reporter Victor Cohn once said that most medical news coverage, particularly front-page coverage, is either "no hope or new hope." What is hopeless this month may be hopeful the next. We need to find the balance in between. Avoid the use of the words "breakthrough" or "cure," however encouraging new findings might be.

3) *Write about the process of science as well as the end results.* Too much of medical coverage is based on ritualistic weekly coverage of the top medical journals, with each study often presented as the latest word on a given risk. We need to step back in covering medicine and present research as an up-and-down, ongoing process that is continually updated as new research comes in. Review articles and features that take the reader inside a laboratory or clinical research setting help put the latest studies in a broader perspective.

4) *Emphasize the degree of uncertainty involved in cancer risks.* Because cancer may develop decades after exposure, the cause and effect is difficult to prove. Too often reporters and researchers try to make certain that which is not. The public can understand that some risks are less defined than others and need more research to help reduce uncertainty and controversy. A risk story should indicate whether research is preliminary or well documented. How does a new study fit in with previous research? How large was the study? How well designed was the study? What reaction is there in the research community to the new findings? What is the magnitude of this risk compared with other risks?

5) *Distinguish between absolute and relative risk.* Too often a new epidemiologic study finds that a given chemical poses a nine times greater risk of causing cancer or other health problems but does not provide the "absolute" risk of getting cancer in the first place. If a risk is 1 in 1 million to begin with, the increased relative risk may be less significant to a given individual than an increase in a more common risk with a baseline of 1 in 1000. Although it did not involve cancer, a good example of the misleading nature of publicizing relative risk alone was a "pill scare" in Britain in late 1995. Regulatory authorities released preliminary findings suggesting that certain new low-dose birth control pills doubled the risk of getting blood clots. The findings were published in the *Lancet* and *British Medical Journal.* Little noticed was a follow-up letter in the *Lancet* noting that the risk of blood clots was so small in the first place that doubling it posed little added danger because it "boils down to whether 9998 or 9997 out of 10 000 pills users remain free" of blood clots *(8).* Journal editors need to ask authors to provide more information on the baseline risk when available.

6) *Distinguish between individual risk and population risk.* Often a given exposure may pose a minimal risk to any single individual. However, if many people are exposed involuntarily to a risk, such as contamination of food, water, or air, it may be a significant public health problem even if the individual risk is small. Stories often confuse individual risk and risk to the population at large.

7) *Stress that exposure is the key, whether the risk is occupational, environmental, or through personal behavior.* There is sometimes a tendency for public health communicators to sound a single alarm bell and try to universalize risk. The early publicity about acquired immunodeficiency syndrome, for example, tended to stress that everyone who was sexually active was at risk. Instead, as the coverage became more sophisticated, there was an emphasis on high-risk populations engaging in certain sexual or drug-use behaviors that greatly increased the likelihood of exposure. Not everyone is at equal risk for any hazard, because the degree and timing of exposure is directly related to increased individual risk. Stories should also stress which populations are most vulnerable to a given risk, such as children, individuals with prior diseases, or older individuals. Too often the alarm bell rings so loudly that everyone feels vulnerable, even when they are not.

8) *Provide information on what can be done about a given risk, whether by the individual, by the public sector, or by government.* It is helpful for the audience to know what the word "prevention" really means and for reporters to distinguish between known preventable risk factors and risk factors that are little understood or uncontrollable. In breast cancer, for example, the strongest known risk factors are family history of the disease in immediate relatives; in the past, there was nothing that could be done about this except increased medical follow-up. Today, the rapid pace of genetic research is providing new opportunities for intervention.

9) *Respect the independence of the news media and recognize the varying needs of different types of journalists.* Like cancer, the word media has a singular sound, yet we too are very diverse. We do not all wake up in the morning and decide on the same story. Much like researchers or clinicians, members of the media are independent practitioners doing their jobs in different ways.

The electronic media, obviously, has different needs than the print media, both in terms of timing and visuals. A daily newspaper reporter has less time to prepare a story and sometimes less space than a reporter working on a long newspaper feature or a magazine writer. The same is true for a daily broadcaster versus a crew working on a primetime "magazine" piece or a documentary. Those who have a story to tell should recognize the needs of individual reporters, as well as our individual autonomy. Most reporters in traditional news media try to avoid being captured by special interests, even those with "good" causes.

10) *Recognize that there is no single "public" that we are trying to reach.* There are many publics out there, who read and listen differently to news of new risks. Some are easily alarmed; others are cynical or convinced that nothing applies to them. Some are literate and well informed; others have little understanding of science or risk. One survey, for example, found that many Americans incorrectly believed that "DNA" stood for Drug and Narcotics Association, a stock market index, or a toxic chemical. Only one of five respondents correctly identified it as a genetic building block or blueprint *(9)*. We need to explain terms clearly and, most importantly, realize that we are frequently preaching to the converted. Often the people we need to reach, who may be at greatest risk, are not the ones reading newspapers or even watching television news on cancer.

Improving coverage of cancer risks is a challenging, long-term project that involves a recognition that there is indeed a problem of often unbalanced, hysterical coverage of cancer and a commitment to improving the product by both those who make the news and those who cover it.

References

(1) Borgman J. Random medical news cartoon. Cincinnati Inquirer and King Features Syndicate 1997 Apr 27; Forum section: 1.

(2) Russell C. The politics of cancer. The Washington Post 1971 Nov 28; Outlook section: C05.

(3) Mankoff R. Mice cartoon. Date unknown.

(4) Kelly W. Pogo cartoon. Cited in Bartlett's Familiar Quotations, 16th ed. Boston: Little, Brown & Co.; 1992. p. 398.

(5) Park WB. Doctor cartoon. The New Yorker 1995 Dec 4; p. 52.

(6) Richter M. Tanker cartoon. The New Yorker 1988 Mar 21; p. 44.

(7) Harris S. Einstein revisited: cartoons on science. New Brunswick (NJ): Rutgers University Press; 1989.

(8) Russell C. The pill is popular but not well understood: new survey shows many women overestimate the risks, underestimate the benefits. The Washington Post 1996 Feb 6; Health section: Z09.

(9) Russell C. How much do people know about health? The Washington Post 1994 Mar 1; Health section: Z06.

Communicating Cancer Risk in Print Journalism

Jane E. Brody

The current barrage of information about real and potential cancer risks has created undue fears and misplaced concerns about cancer hazards faced by Americans. Most members of the general public are far more worried about minuscule, hypothetical risks presented by environmental contaminants than about the far greater well-established hazards that they inflict on themselves, for example, through smoking, dietary imbalance, and inactivity. It is the job of the print media to help set the record straight and to help place in perspective the myriad cancer risks that are aired almost weekly in 30-second radio and television broadcasts. [Monogr Natl Cancer Inst 1999;25:170–2]

Health reporting to the public has changed dramatically in recent decades.

- Whereas doctors were once terrified of speaking to reporters, today, many hire public relations agents to help them get their names in the news. The journals, too, compete for media attention—and they get it, especially when their reports involve cancer.
- Whereas medical researchers were once reluctant to speak beyond the direct implications of new findings, today, even when not pressed by reporters, they often extrapolate freely and speculate wildly about far-reaching consequences.
- Whereas it once took a dramatic treatment, a new cure, or the discovery of a major health hazard to prompt an editor to make space for a medical news story, now every minor little development is likely to gain a public airing in print and broadcast media throughout the country.
- Reporters, editors, and publishers now recognize an intense public interest in health news and a seemingly insatiable appetite for what I call "health hype"—the latest proclaimed cancer-fighting food or nutrient; the latest health threat, real or imagined, in our food, water, or air, especially if the threat involves cancer; the latest discovery, real or imagined, of a cancer cluster, and so forth.

The public is now suffering from a serious case of information overload. People are constantly being bombarded with information they are ill equipped to process. This fact is especially true of brief reports aired on radio and television. It is not possible in 30, 60, or even 90 seconds of air time to describe all the ifs, ands, and buts that a new report on cancer usually warrants and often requires if listeners are to understand its implications and to appreciate its relevance—or lack thereof—to their personal lives. And, to make matters worse, radio and television news reports are often prepared by people with little or no specific knowledge of medical science. The reporters lack a context and a critical eye and all-too-often produce reports that are seriously distorted if not downright incorrect.

The public has naturally come to view the cancer risks most talked about as those that are most serious. That is why one

Correspondence to: Jane E. Brody, M.S., Science News, *The New York Times*, 229 West 43 St., New York, NY 10036.

© Oxford University Press

quarter of the population, according to a CBS News Poll last year, believes that breast cancer is the leading cause of death in American women. That is also why breast cancer research receives about $600 million a year from the federal government, whereas lung cancer, which each year kills nearly four times more Americans overall and 50% more American women than does breast cancer, receives only about $100 million from the federal money pot. It is an example of the squeaky wheel phenomenon, with the noise in part generated and frequently enhanced by a constant barrage of media reports on the disease American women have come to fear most.

The changes in health communications outlined above have added up to massive public confusion about the causes of cancer and how some of the leading kinds of cancer might be prevented or treated. It never ceases to amaze me how a risk as minuscule as the one attributed to the growth regulator Alar®, for example, could have generated such mass hysteria—to the point where at least one mother raced after her son's school bus to remove the apple from his lunch pail—while countless mothers, many of whom were among those who poured millions of gallons of apple juice down the drain, continued to smoke cigarettes, a well-known and far more potent cancer killer than Alar was or ever might have become *(1)*. The U.S. Environmental Protection Agency, which banned Alar, estimated that lifetime exposure to Alar would increase a person's risk of dying of cancer by 45 ten-thousandths of 1%—0.000045. That risk pales by comparison to the known hazards of cigarette smoking that multiplies cancer risk by 20 or more times, not to mention the potentially fatal damage it inflicts on the heart and blood vessels.

The problem with communicating about cancer risks is that people have become so fearful of cancer that anything at all said about it tends to be viewed in absolute, not relative, terms. A "possibility" becomes a probability or even a certainty in the public mind, especially when the factor involved is one that is inflicted on us by industry or government. The far more serious risks that are self-inflicted—like cigarette smoking; a low-fiber, high-fat diet; obesity; inactivity; and so forth—do not invoke nearly as much concern or action.

As I see the job of the print media, it is to bring perspective and depth to news reports about cancer. Although I do not suggest that we ignore reports that are likely to get a wide hearing in the mass media, I do suggest that the print media must always be certain to take the time and to allow the space needed to place the report in a meaningful and accurate context. If the matter is of sufficient interest and importance to the public, it can and should be dealt with in a second-day story that adds breadth and depth to a new finding, which may be difficult to do the first day before outside experts have had a chance to review the original report.

I recall a report published in 1981 in the *New England Journal of Medicine* by researchers at Harvard *(2)*. It linked coffee drinking to pancreatic cancer, a particularly deadly disease, the possible causes of which little is known beyond cigarette smoking. The story aired on every broadcast and was published in every newspaper in the country, including page 1 of *The New York Times*, even though I had raised serious questions about the validity of the finding based on what I knew of laboratory studies of coffee and caffeine in relation to cancer and of the well-established association between coffee consumption and cigarette smoking.

In subsequent reports, several major epidemiologic studies, including the American Cancer Society's million persons study, weighed in. All reported finding no hint of an association between the consumption of coffee, caffeine, or decaffeinated coffee and pancreatic cancer (or any other cancer, for that matter). But what stuck in the public's mind is that coffee may cause a deadly cancer, a notion that was not even dislodged when the Harvard group repeated its study 5 years later and failed to confirm its original findings *(3)*.

It is more than trivial interest, however, to review the consequences of the initial Harvard report. It sparked virtually no change in coffee consumption beyond a small decline for a day or two after the initial report. Whereas the Alar scare nearly killed the apple industry in the United States and bankrupted many growers, the coffee industry never even felt a pinch. You see, coffee drinking is a self-chosen "risk"—one near and dear to American hearts, taste buds, and nervous systems—whereas Alar was being applied to apples by people other than consumers, who had little or no appreciation for its value to producers and did not really care that it benefited anyone else if it might be a hazard to them.

Americans are even willing to fight to maintain their exposure to potential hazards if they consider them to be valuable assets. You no doubt recall the public clamor to keep saccharin on the market after it was declared a carcinogen by the U.S. Food and Drug Administration. Whether saccharin is dangerous or not is not the point. The point is that the public is not capable of rationally considering risks, and the media are at fault for perpetuating this ineptitude.

A trend in print reporting in recent years offers some hope for improving the public's understanding of risks and the ability to more rationally weigh risks versus benefits. Dozens of health newsletters are now being published, including quite a few that offer advice based on solid medical research. These health newsletters do not depend on flashy headlines and hot-off-the-press news to attract readers. They are not afraid to say when evidence is shaky, inconclusive, or not worth the paper it is printed on and, even more important, to explain why a finding should or should not be ignored.

In newspapers too, a clear trend is moving toward more in-depth reporting; toward a tendency to wait days, weeks or even months to put new findings in their proper perspective; and toward considering different points of view, based on solid research, not hearsay. In fact, many reporters who may not have access to primary journals or the ability to digest them accurately can now use secondary sources like these newsletters as well as the throw-away medical magazines for background and context on an unlimited range of health topics, including cancer.

Let me give you a good example of responsible reporting in a health newsletter. In 1989, Harvard researchers reported in the journal *Lancet (4)* a finding that women who developed ovarian cancer reported having eaten more yogurt over the previous 5 years than did healthy women. You can hear the 15-second radio news report: "Eating yogurt has been linked to deadly ovarian cancer, according to researchers at Harvard University, one of the nation's leading medical institutions."

Well, here is how the *Nutrition Action Healthletter*, published by the Center for Science in the Public Interest, described the finding *(5)*: "Finding a culprit [for ovarian cancer] would be terrific, but it's too early to blame yogurt." The newsletter went on to explain that the researchers speculated that the association with yogurt may be as a result of an inability to process galac-

tose, one of the sugars in yogurt, because of an enzyme deficiency. But the newsletter noted that how the enzyme might be related to cancer is unknown, that the cancer might have caused this enzyme deficiency rather than the other way around, or that the entire finding might have been a statistical fluke. Then it quoted the researchers as stating that "further human epidemiological studies and relevant animal experiments are warranted before any public health recommendations can be made." The Healthletter report concluded, "We agree."

Now, that is responsible reporting of a possible cancer risk, a report that should not prompt women to abandon an otherwise healthful food on the basis of highly preliminary evidence and a report that helps readers understand what is behind the new finding and what needs to be done to confirm or refute it before anyone changes her diet.

As the Healthletter accurately pointed out, "Many inconclusive or poorly designed studies get published—and quoted. While the researchers often acknowledge the limitations of their work, the message can get lost in the popular press"*(5)*.

The problem with much cancer risk reporting is that the new and surprising but relatively minor findings are likely to get far more attention than the better established and far more serious ones. Readers shrug when yet another report links cigarette smoking or a high-fat diet to an increased risk of a common, deadly cancer, and they go right on puffing away and smearing butter on their bread and chomping on 12-ounce well-marbled steaks. Just look at what has happened to cigars. After 3 decades of decline in cigarette smoking, especially by American men, cigars have emerged as the hall mark of sophistication, with special bars, magazines, and stores to foster this noxious trend.

Yet if I wrote a story stating that living for 20 years near a polyvinyl chloride plant increases the risk of death by one part per million (or, to put it another way, reduces life expectancy by 8 minutes), chances are a lay group would take up a cause to banish such plants or have them encased in concrete. And what do you think would be the consequences of a story stating that residues of the now-banned pesticide ethylene dibromide (EDB) incurred the same potential cancer risk as eating two raw mushrooms a day? I can tell you what would happen, because I wrote that story and it did happen. Many of my readers cried, "Now I can't eat mushrooms"!

As you already know, the public is much more likely to become exercised about an involuntary risk than a voluntary one, an unfamiliar substance than a familiar one, a new hazard than a much greater, but well-known one. However, if reporters are to fulfill their responsibility to the public—which is to inform, not mislead—we have an obligation to continue to harp away on what is really important and pay far less attention to minor risks—minor either in terms of the absolute risk or in the numbers of people at risk. We have an obligation to remind women over and over again that it is not the radiation from a mammogram they should be fearing, but of not getting a mammogram and failing to detect a breast cancer in its early, most curable stages.

We have an obligation to report every new finding that links fruits and vegetables and dietary fiber to protection against cancer, as well as to carefully and objectively examine the evidence for or against a relationship between various kinds of dietary fats and cancer.

But most of all, we have an obligation to keep the most important cancer risk—tobacco—squarely in public view. I would even suggest that every newly reported cancer risk be compared with the risk of smoking, because I am willing to bet that it will pale by comparison. And in doing so, we will continue to remind people to focus on what is most important and what they are most able to control—their own living habits.

References

(1) Marshall E. A is for apple, Alar, and . . . alarmist? Science 1991;254:20–2.

(2) MacMahon B, Yen S, Trichopulous D, Warren K, Nardi G. Coffee and cancer of the pancreas. N Engl J Med 1981;304:630–3.

(3) Hseih CC, MacMahon B, Yen S, Trichopoulus D, Warren K, Nardi G. Coffee and pancreatic cancer. N Engl J Med 1986;31:587–9.

(4) Cramer DW, Harlow BL, Willett WC, Welch WR, Bell DA, Scully RE, et al. Galactose consumption and metabolism in relation to the risk of ovarian cancer. Lancet 1989;2:66–71.

(5) Center for Science in the Public Interest. Nutrition Action Healthletter. Washington (DC): The Center; October 1989. p. 4.

Challenges to Improving Health Risk Communication in the 21st Century: a Discussion

Geoffrey T. Fong, Lynn A. Rempel, Peter A. Hall

Risk communication has always been fraught with challenges, and there is perhaps no greater example of the range and complexity of the challenges involved in risk communication than in the domain of health risk communication and, more specifically, in the domain of cancer risk communication. In this paper, we discuss some important issues that we believe should be addressed to meet the many challenges that face researchers, health communicators, and the media in communicating health risks to patients, to their families, and to the public.

How Should We (or Should We) Communicate the Uncertainty Associated With Risk Information?

Uncertainty is the bane of a risk communicator's existence, and this is particularly true for those who communicate risk in a domain in which the knowledge base is changing rapidly, such as cancer risk: Not only is the risk estimate expressed in terms of uncertainty, such as a probability, but also the risk estimate itself is subject to uncertainty. Consider, for example, the wide range of estimates for lifetime risk of ovarian cancer among women with BRCA1 mutations. The estimate from one study of BRCA1-linked families *(1)* was 30%, but with a 95% confidence interval of 8%–47%; the estimate from a population-based study *(2)* was 16%, with a 95% confidence interval of 6%–28%. A commonly cited estimate of lifetime risk is 50% [e.g., *(3)*], and a recent text *(4)* summarizes the literature by suggesting that the lifetime risk is somewhere between 10% and 80%! Given the inherent challenges of generating valid population risk estimates from either cancer registry families or special populations (e.g., Ashkenazi Jews), along with the instability in estimates derived from small samples, it is not surprising that there is considerable uncertainty in risk estimates for many hereditary cancers.

But how then should the risk communicator convey this risk information to the individual (or to the public)? Does one present the estimate from the single, presumably most relevant or highest quality study, if that is even ascertainable, or does one present an average of estimates? As Schwartz et al. *(5)* point out, communicating a single risk estimate without uncertainty implies a greater degree of precision than is warranted, and they advocate using confidence intervals to capture this second-order uncertainty, also known as "ambiguous" or "vague" probabilities *(6)*. However, presenting a range of risk estimates, rather than a single risk estimate, creates two problems. The technical problem is that it is unclear which range should be presented. Should we present the range of mean estimates across studies (e.g., "16%–30%") or the range of confidence intervals across studies (e.g., "6%–47%")? The communication problem is that presenting a range of risk estimates may evoke negative reactions of confusion and lower trust *(7)*, leading recipients to feel that they have not received sufficient information to assist them in their decision process *(8)*. Telling a woman that her risk of ovarian cancer is between "10% and 80%" seems to say, "Our state of knowledge is so poor that we really don't know what your risk is."

We believe that this is the basic dilemma facing the health risk communicator—one that has both ethical and pragmatic aspects: a conflict between our desire to accurately convey the limits on our present knowledge, which pushes us in the direction of presenting a range of risk estimates, and our desire to provide risk information that empowers recipients to make confident risk-relevant decisions, which pushes us in the direction of presenting a single "best" estimate. How this conflict is resolved will have a powerful effect on an individual's or the public's reactions and will have implications for every aspect of the risk communication itself. For example, the complexity associated with multiple estimates might well make graphical presentations unwieldy and confusing (imagine representing confidence intervals with a series of pie charts). There is a need for research on the effects of vague probabilities on the judgments and decisions about risk and on recipients' trust and confidence in the risk information and in the risk communicator. Such research could help guide those critical initial decisions about how to present risk information to patients and to the public.

What Format Is Most Effective in Communicating Risk Information?

Lipkus and Hollands *(9)* provide an excellent review of the existing literature on the effects of visual or graphical displays of risk information. Although research on the cognitive processing of graphs [e.g., *(10–12)*] has been applied to understanding comprehension and responses to graphical presentations of risk [e.g., *(13,14)*], we still know little about what graphical methods could lead to more effective health risk communications. In fact, we know little about how (or even whether) graphical presentations are superior to other formats, such as numeric or verbal formats. Because graphical presentations can be so powerful, it is particularly important for research to discern the best methods for presenting information—methods and formats that will maximize comprehension of the message. In short, if a picture is worth a thousand words, we need to ensure that it is conveying the appropriate words.

How and Why Does Tailoring Work?

Recent advances in new interactive technologies now make it possible to create powerful and sophisticated methods for risk assessment, for risk communication, and for delivering interventions to reduce health risk. The review by Strecher et al. *(15)*

Affiliation of authors: Department of Psychology, University of Waterloo, Ontario, Canada.

Correspondence to: Geoffrey T. Fong, Ph.D., Department of Psychology, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: gfong@watarts.uwaterloo.ca).

See "Notes" following "References."

© Oxford University Press

illustrates so clearly that these technologies represent the engine that will revolutionize the nature of risk communication and health interventions, in research as well as in practice.

These new interactive technologies allow the construction of tailored interventions that are much more fine grained and elaborate than has been possible in the past. The review of the literature by Rimer and Glassman *(16)* provides fairly good evidence for the effectiveness of tailoring, although many of the studies cited do not demonstrate the beneficial effects of tailoring per se because they confound tailoring with other aspects of the intervention (e.g., quantity and specificity of the information). Rimer and Glassman also provide a nice classification scheme for thinking about the dimensions on which tailoring could be accomplished and offer an excellent set of suggestions for future research on tailored risk communications. We wish to amplify one of their points and add another. First, it is important for research to identify which of the many variables that could be employed in a tailored intervention will actually lead to increased effectiveness. Interactive technologies have given us methods for tailoring on virtually any dimension, and now we must engage in the hard work of finding out which variables are worth tailoring on and whether the benefits outweigh the costs. If this interactive technology is the engine, well-designed research should be the navigator that will guide us toward the creation of more effective interventions.

The second point, which follows from the first, is that it is important for us to identify those theoretical models that are best suited for guiding the creation of tailored interventions. One of the attractive features of stage theories, such as the transtheoretical model *(17,18)* and precaution adoption theory *(19)*, is the assumption—sometimes explicit and sometimes implicit—that interventions that are tailored to an individual's stage of change will lead to greater behavior change. The proper research design for testing this hypothesis is to compare the effects of a stage-appropriate intervention with those of a stage-inappropriate intervention. For example, to properly test the matching hypothesis in the transtheoretical model, an intervention that is tailored for contemplators should be given not only to contemplators but also to individuals at other stages (e.g., preparers). Such fully crossed factorial designs (i.e., those that cross stage of intervention with stage of participant) have not typically been employed [for a notable exception, *see (19)*]; thus, there still remains a surprising paucity of empirical evidence for the superiority of matching intervention to behavioral change stage, despite the strong intuitive appeal of that hypothesis.

Rimer and Glassman *(16)* urge researchers to employ factorial designs to identify components of a risk communication that may be most effective. In addition, we urge researchers to measure the theoretically relevant mediators of behavior change so that we can understand how and why a risk communication is effective. In the transtheoretical model, for example, the processes of change are theorized to be the psychological substrate from which the individual progresses from one stage of change to the next. If so, the effects of a matched intervention on behavior change (or progression from one stage to another) should be accompanied by effects on the processes of change associated with the current stage and the next stage. Measuring these mediators within a longitudinal design would allow researchers to perform a rigorous empirical test not only of the intervention but also of the theoretical assumptions of the model *(20,21)*. In a recent example of this methodological strategy, Wang et al. *(22)* examined the effects of a genetic risk assessment among 60 self-referred patients at an ovarian cancer clinic in a three-wave longitudinal study in which the outcome variable was intentions to engage in future screening behaviors at 2-day follow-up. They found that risk of ovarian cancer that was communicated to patients during the genetic counseling session was significantly related to future screening intentions, and mediational analyses suggested that this effect was because of changes in postsession perceived risk and not because of changes in postsession cancer worry. These results suggest that perceived risk may play a central role in the pathway from risk communication to risk-relevant behavior.

Can Researchers and Journalists Work Together to Improve Health Risk Communication to the Public?

There has been a tremendous increase in health reporting in recent years. Stories about health are being given more space in the print media and more time in the electronic media. The popularity of health reporting has led to an amplification of both the negative and the positive aspects of health reporting. On the negative side is that some health reporting has followed the same seemingly inexorable descent into sensationalism that has characterized reporting in many other domains. On the positive side is the trend toward more in-depth reporting on important health risks, at least among the "elite media" *(23)*. Both Brody *(24)* and Levine *(25)* point out that the public has considerable misapprehension about health risk information, although they focus on different aspects of the problem. Brody provides compelling testimony to the dramatic changes in health reporting and in the public's awareness of health risks that have occurred recently. Her concern that the public is suffering from information overload is well-taken; unfortunately, the overload problem will only become greater in the future, and it is unclear how or whether it can be ameliorated.

Levine *(25)* focuses on the professional, cultural, and epistemic differences between researchers and journalists and suggests that these differences have created a "chasm" between the two. Indeed, there are a number of differences between the two professions that may conflict. First, differences exist in time frame. Journalists have considerably shorter deadlines and very limited space to tell their story. They have neither the time nor the space to describe the caveats and the nuances of the research. Second, differences exist in epistemic goals and standards, which are nicely summarized by a statement that a journalism professor made to the first author years ago: "Reporters want simple definite statements. They can't stand two-handed researchers—those who say, 'on the one hand this . . . on the other hand that. . . .' " But as we know, the researcher's métier is, in fact, "two-handed thinking," e.g., exploring the boundary conditions for a phenomenon, identifying the conditions under which its opposite may obtain *(26)*, or considering alternative explanations. This conflict between the journalist's need to communicate certainty and the researcher's need to acknowledge the uncertainty that is a part of any research endeavor is analogous to the dilemma, described earlier, that is faced by the risk communicator in deciding whether or not to acknowledge the uncertainty of a risk estimate.

Can researchers and journalists work together more effectively to improve media coverage of health risks? We believe that this is possible; however, given the more severe structura

constraints imposed on journalists, we suspect that it is the researchers who need to change to a greater extent. Researchers can inerease their awareness of the broader implications of their research; they can work together with their public relations office so that press releases describing their findings can be crafted in ways that fairly negotiate between the opposing goals of precision and simplicity. Professional organizations and research funding agencies can offer assistance in the form of workshops and guidelines for communicating with the media.

How Can We Best Meet the Future Challenges Facing Health Risk Communication?

How will the public obtain information about their health risk in the future? Rapid advances in molecular biology, for example, biochip technology *(27),* will make it considerably easier and less expensive for individuals to obtain information about their genetic risk. As with most technological advances, this will lead to an amplification of both the benefits and the costs. For example, it is doubtful that the safeguards for communicating genetic risk information currently embodied in informed consent and other guidelines will be honored or protected when the risk communication emanates from sources outside the medical research community; thus, the complex legal and ethical issues surrounding genetic testing will only become more problematic.

We are currently undergoing a revolution in how information is disseminated or transmitted to the public by the media (indeed, the very concept of mass media is being transformed). This revolution will lead to dramatic changes in how the public obtains information about health risks. Even today, the vast majority of the public obtains its information about health risks not from the elite media, such as the *New York Times* or *The Washington Post*, but from popular television programs or from other non-elite print and electronic sources. What will be the effect of this shift from more controlled, vertical, and hierarchical communieation transmission channels to more uncontrolled, horizontal, and nonhierarchical ones? Studies of comprehension of information from physicians suggest that, even under the best of circumstances, a significant proportion of information that is conveyed to patients is either not remembered, misremembered, or misinterpreted. If this is the case, health information that is obtained from the media may be subject to even greater distortion. And looking to the future, with the exponential growth of the Internet, there is a huge and growing number of health-related web pages whose information is of uncertain or dubious validity *(28,29).* Efforts to create standards for interactive health information *(30)* are much needed and well intentioned, but the ease with which anyone can create a vivid and compelling web site where health risk information is disseminated without any regard to its validity suggests to us that such standards may ultimately have less impact than they should on ensuring the validity of health risk information on the Internet.

What, then, can be done to counteract these forces that threaten the validity of health-risk information? We have three broad suggestions. First, given the impossibility (and possible unconstitutionality) of regulating content on the Internet and other media, it is necessary instead to fight fire with fire. This can take several forms. For example, institutions such as the National Institutes of Health, the Centers for Disease Control and Prevention, and other governmental entities must endeavor to enhance their reputations as trusted and reliable sources of information about health risk by being proactive. They should monitor the public's concerns about any given health risk, then quickly inform the public about the validity of that health risk, and offer guidelines or suggestions based on the best available sources about the seriousness of the risk and about the methods of reducing the risk, if the risk is serious. In the same way that some corporations have sizable public relations staff that monitor the media for relevant news and provide rapid response for any media report that is worthy of response, so too could agencies provide rapid response for media reports about health risks that are either worthy of amplification or clarification or in need of rejection. Recently, there have been vigorous and concerted efforts by institutions such as the National Cancer Institute to become a more proactive source for health risk information; this is an admirable and important development.

Second, it would be beneficial for guidelines to be created for improving the public's understanding of health risk. Recently, an advisory group convened by the Harvard School of Public Health and the International Food Information Council Foundation issued such guidelines for communicating scientific research on nutrition, food safety, and health. That statement included communication guidelines for scientists; journal editors; journalists; and industry, consumer, and other interest groups, as well as some general guidelines for all parties *(31).* Creating a similar set of guidelines in the domain of communication of cancer risk, or health risks more generally, might well be beneficial for improving both the communication and comprehension of such risks.

Third, it is important to educate the public about the basics of probability theory and statistics—the language of risk. We concur with the recommendations by Schwartz et al. *(5)* to develop methods for educating patients so that they have a better understanding of quantitative concepts, such as probability. H. G. Wells's belief that "statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write" has never been so relevant as it is today and will be in the future. Enhancing the public's understanding of probability and statistics would involve additions to school curricula, but other sources could contribute as well. Recent popular books on innumeracy *(32)* have heightened the public's awareness of the importance of understanding quantitative concepts, and other such efforts are urgently needed.

In summary, we are confident that research will lead to methods for communicating health risks that maximize comprehension and facilitate informed health judgments, decisions, and risk-consistent behavior within controlled settings, such as in genetic counseling or health promotion programs. However, we are less sanguine about whether the same advances can be expected in the uncontrolled environment of the mass media, including the Internet. We fear that much information about health risk that is encountered by the public may be untrue, partially true, or misinterpreted. Incorrect or misleading information may even be deliberately generated by advocates who have an ulterior motive for their actions *(28,29).* It will take considerable efforts to counteract various forces that threaten the integrity of health risk information.

The stakes are high. We hope that the work of researchers, health risk communicators, and journalists alike will increase our understanding of the factors that contribute to effective risk communication. In so doing, we will be taking important steps toward increasing behaviors and formulating policies that may reduce our health risk in the 21st century.

References

(1) Easton D, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. The Breast Cancer Linkage Consortium. Am J Hum Genet 1995;56:265–71.

(2) Struewing J, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, et al. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 1997;336:1401–8.

(3) Mange E, Mange AP. Basic human genetics. Sunderland (MA): Sinauer Associates, Inc.; 1998.

(4) Offit K. Clinical cancer genetics: risk counseling and management. New York (NY): Wiley; 1998.

(5) Schwartz LM, Woloshin S, Welch HG. Risk communication in clinical practice: putting cancer in context. Monogr Natl Cancer Inst 1999;25: 124–33.

(6) Budescu D, Weinberg S, Wallsten TS. Decisions based on numerically and verbally expressed uncertainties. J Exp Psychol Hum Percept Perform 1988;14:281–94.

(7) Johnson B, Slovic P. Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. Risk Anal 1995; 15:485–94.

(8) Fox C, Tversky A. Ambiguity aversion and comparative ignorance. Q J Econ 1995;110:585–603.

(9) Lipkus IM, Hollands JG. The visual communication of risk. Monogr Natl Cancer Inst 1999;25:149–63.

(10) Cleveland W, McGill R. Graphical perception: theory, experimentation, and application to the development of graphic methods. J Am Stat Assoc 1984;70:531–54.

(11) Hollands J, Spence I. Judging proportion with graphs: the summation model. Appl Cogn Psychol 1998;12:173–90.

(12) Simkin D, Hastie R. An information processing analysis of graph perception. J Am Stat Assoc 1987;82:454–65.

(13) Stone E, Yates JF, Parker AM. Risk communication: absolute versus relative expressions of low-probability risks. Org Behav Hum Decis Process 1994;60:387–408.

(14) Stone E, Yates JF, Parker AM. Effect of numerical and graphical displays on professed risk-taking behavior. J Exp Psychol Appl 1997;3:243–56.

(15) Strecher VJ, Greenwood T, Wang C. Interactive multimedia and risk communication. Monogr Natl Cancer Inst 1999;25:134–9.

(16) Rimer BK, Glassman B. Is there a use for tailored print communications in cancer risk communication? Monogr Natl Cancer Inst 1999;25:140–8.

(17) Prochaska J, DiClemente CC. Stages and processes of self-change in smoking: toward an integrative model of change. J Consult Clin Psychol 1983; 51:390–5.

(18) Prochaska J, DiClemente CC, Velicer WF, Rossi JS. Standardized, individualized, interactive, and personalized self-help programs for stages of smoking cessation. Health Psychol 1993;12:399–405.

(19) Weinstein N, Lyon JE, Sandman PM, Cuite CL. Experimental evidence for stages of health behavior change: the precaution adoption process mode applied to home radon testing. Health Psychol 1998;17:445–53.

(20) Fong GT, Jemmott JB III, Jemmott LS. The theory of planned behavior a a mediational model for an HIV risk-reduction intervention for Africa American adolescents. Proceedings of the Society of Behavioral Medi cine's 20th Annual Meeting; 1999 Mar 2–6; San Diego (CA). Ann Beha Med 1999; 21(Suppl): S200–1.

(21) West S, Aiken LS. Toward understanding individual effects in multicom ponent prevention programs: design and analysis strategies. In: Bryant K Windle M, West SG, editors. The science of prevention: methodologica advances from alcohol and substance abuse research. Washington (DC American Psychological Association; 1997. p. 167–209.

(22) Wang C, Fong GT, Ritvo PG. The psychological effects of genetic ris communication: a longitudinal study. Proceedings of the American Psy chological Association; 1998 Aug; San Francisco (CA).

(23) Klaidman S. How well the media report health risk. Daedalus 1990;119 119–32.

(24) Brody JE. Communicating cancer risk in print journalism. Monogr Na Cancer Inst 1999;25:170–2.

(25) Levine J. Risky business—communicating scientific findings to the public Monogr Natl Cancer Inst 1999;25:163–6.

(26) Salovey P, Schneider TR, Apanovitch AM. Persuasian for the purpose c cancer risk reduction: a discussion. Monogr Natl Cancer Inst 1999;25 119–22.

(27) Chee M, Yang R, Hubbell E, Bemo A, Huang XC, Stern D, et al. Accessin genetic information with high-density DNA arrays. Science 1996;274 610–4.

(28) Bower H. Internet sees growth of unverified health claims. BMJ 1996;313 381.

(29) Keoun B. Cancer patients find quackery on the Web. J Natl Cancer Ins 1996;88:1263–5.

(30) Robinson T, Patrick K, Eng TR, Gustafson D, and the Science Panel o Interactive Communication and Health. An evidence-based approach t interactive health communication. JAMA 1998;280:1264–9.

(31) Improving public understanding: guidelines for communicating emergin science on nutrition, food safety, and health. J Natl Cancer Inst 1998;90 194–9.

(32) Paulos J. Innumeracy: mathematical illiteracy and its consequences. Nev York (NY): Vintage Books; 1990.

Notes

Supported by grants from the Ontario Cancer Genetics Network and the Socia Science and Humanities Research Council of Canada to Dr. Fong. L. A. Rempe and P. Hall were supported by graduate fellowships from the Social Science an Humanities Research Council of Canada.

We thank Cathleen McDonald for her comments on a preliminary draft an Alexandra Fong for her assistance.

Implications for Improving Risk Communication Through Various Channels: a Discussion

Cristine Russell

The news media need to improve their coverage of new and old cancer risks, avoiding sensational reporting of minor risks and underreporting of major ones. However, efforts to communicate about cancer risk to the general public need to go far beyond the traditional print and broadcast outlets into more innovative, personalized channels that can better educate individuals about the risks they face and what they can do about them. [Monogr Natl Cancer Inst 1999;25:177–8]

The session, "Implications for Improving Risk Communication Through Various Channels," offered an overview of the multiple opportunities for improving the public's understanding of health risks and taking action to minimize these risks. One of the most surprising aspects of this session, and indeed of the conference in general, was the relative lack of "media bashing." As a journalist, I find it refreshing not to have the news media blamed wholly for the problem—that is, the lay public's lack of understanding, as well as downright misunderstanding, of health risks—or to have the media held accountable for solving the problem.

In fact, as was clear in these presentations, improving public understanding of risk is a shared problem and a shared opportunity that involves a broad array of players, including those in traditional print and broadcasting news, entertainment, advertising, public health, health communication, clinical medicine, scientific research, government at all levels, and professional and lay advocacy groups.

The "target" is, of course, the "public." However, one important acknowledgment in all of the presentations was the obvious but often ignored fact that there is no single "public," particularly when it comes to communicating about the health risks faced by individuals and communities at large. There are many publics, not only in terms of sex, age, and geography, but also in terms of individual risk susceptibility, exposure, and risk literacy. Often those at greatest potential risk are least likely to be well informed about the risks they face or able to understand the complex, changing data about health hazards. Conversely, many of the most voracious consumers of health information through the traditional mass media channels of newspapers, magazines, television news, and books are a more upscale, literate audience. They may be better informed but at lower risk because they are already taking good care of their health and have greater access to the health care system. To make a dent in the difficult task of improving health risk understanding across the board, one must address the many "publics" in the real world.

In this regard, a two-pronged approach is important. Health risk information will continue to be transmitted in the traditional mode through the mass media. At the same time, considerable effort must be made to provide more opportunities for concrete, highly directed, high-quality risk messages to reach targeted audiences about prevention and medical decision making.

In the first instance, there is no doubt that tremendous improvement should be made in the way health risk information is transmitted to the public through the mass media. Over the past two decades, the sheer amount of risk information on everything from saccharin in soft drinks to asbestos in schools has contributed to a somewhat hysterical view of risk, particularly a public feeling that "everything" indeed seems to cause cancer. To some degree this is inevitable. The news media seldom cover studies showing that something does not cause cancer. News coverage by definition involves a focus on new rather than familiar risks, on conflict, on uncertainty, and on human drama. These characteristics are seen more often in environmental, political, or legal arenas than in doctors' offices or scientists' laboratories.

Broadcast journalist Jeff Levine correctly noted in his paper that there is a natural "chasm between the scientific worldview and the journalistic approach to handling information. Risk to the scientist is ultimately a tool to measure and compare data. For journalists, risk is a measurement of news value. The greater the risk of anything, the bigger the story. . . . Statistics convince scientists of risk, but they don't sell papers or get people to watch television newscasts." Levine says improvements in risk communication depend on both scientists and journalists. Scientists need "to know how to explain, within limits, what are the risks and benefits of a given problem in plain English." Journalists, meanwhile, need to be "more restrained and cautious," qualities that he notes are difficult to come by "in an environment soaked with sensation of every sort." Strategies for getting scientists to be better communicators and journalists to be less sensational he leaves to others, but the goals advocated by Levine are nonetheless worthy ones.

In her paper, *New York Times* science writer Jane Brody stressed that the responsibility of the news media is "to inform, not mislead." She chastised her colleagues for overemphasizing minor risks and said journalists "have an obligation to keep the most important cancer risk—tobacco—squarely in public view." Actually, in recent years, tobacco risks have received more coverage, but smoking rates among young people continue to climb. Negative news coverage alone will not eradicate tobacco use.

Journalistic coverage of cancer risks can indeed improve with more training, experience, and understanding of science and medicine among both reporters and editors. But efforts to reach the public and attempts to change behavior need to go far beyond the news media.

As this session demonstrated, efforts are already under way by researchers and practitioners seeking innovative ways to transmit risk information to the public. In the advertising world, there is great potential for going far beyond mass media public service announcements, which in the past have often seemed

Affiliation of author: Freelance writer and special health correspondent, *The Washington Post.*

Correspondence to: Cristine Russell, 43 Huckleberry Ln., Darien, CT 06820.

© Oxford University Press

dreary, out of date, and of questionable public impact. In these presentations, one learned that there is an encouraging proliferation of research into alternative approaches. One of the most promising is the effort to reach "niche" audiences with far more personalized, and, thereby, more relevant, information about the individual health risks they face and what they can do about them. It is a little bit like the expansion of television options from three main networks into myriad cable channels, in essence going from "broad band to narrow band" to cater to special interests.

Also promising are new initiatives to go into the physician's office to provide personalized information to individual patients. It has long been recognized that physicians lack the time to really talk with patients about their personal health issues and options. The inference has been that somehow time is the main problem, and, if only physicians had more of this precious commodity, they could in fact communicate with their patients about the complex issues of medicine. I am not convinced that this is in fact the case. It is not clear that physicians are any better equipped than the rest of us to explain complex data, particularly risk data, to their individual patients. Communication training is seldom part of medical school or continuing education curricula. Physicians often rely on anecdotal information and in many cases do not believe the epidemiologic studies that contradict their own experience. Therefore, part of the problem is not just translating new and complicated health risk information for the public or the mass media but training physicians, nurses, and other health professionals to talk with patients.

Dartmouth Medical School physicians Lisa M. Schwartz and Steven Woloshin are leaders in this effort to improve health risk communication in physicians' offices or clinical settings. They recommend further development of simple office-based tools to help personalize risk information for individual patients using "low-tech" charts that display risks of dying of various diseases and the possible health benefits of changing behavior (e.g., stopping smoking) or seeking regular medical screening (e.g., mammograms or Pap tests for women). Whether physicians and patients will use such approaches remains to be seen, but I agree with Schwartz and Woloshin that it is certainly worth a try.

A different approach to "mass customization" was advocated by Barbara K. Rimer of the National Cancer Institute. She described the use of "tailored print communications" to provide cancer risk information for a given individual. Advances in computer technology and genetics studies enhance the prospects for developing newer approaches, with better risk information, than was available in the past. Again, however, one of the bigges obstacles is getting risk information into the hands of the peopl who need it most. In the past, individual health risk assessment seemed promising for widespread use but have instead bee limited largely to settings such as the workplace.

The paper by Michigan health communication researche Victor J. Strecher and colleagues suggested some inventive ne ways of getting far more effective risk information to individua users through "interactive multimedia," such as the Internet interactive CD-ROMs, and computer kiosks. One danger is tha the Internet and CD-ROMs tend to self-select again for a mor educated, affluent audience. I liked the idea of using kiosks i shopping malls, supermarkets, community centers, and othe "high traffic" areas to reach out to an audience that is younge and more ethnically and economically diverse. This approac could expand the base of health literacy to a wider spectrum o people who may not have as much access to computers. Is i possible that "risk" presentations could actually be fun *and* po tentially influence behavior? Strecher's presentation suggest that this could indeed be the case (could MTV risk communi cation be far behind?).

Ultimately, whatever the medium, there is still the problem o controlling both the quantity and the quality of risk informatior It is a challenge to keep risk information up to date and accurat and to avoid information overload. Bombarding the public wit too much information, however helpful, will probably backfire It is also important to carry out research to see if getting bette health risk information truly makes a difference in health be havior.

Finally, I would urge health risk communication enthusiast to avoid the word "persuasion." It is important to be cautious i terms of having all the answers and adopting an almost religiou approach to converting consumers or patients to the "right way of behaving or acting. Rather than attempting to persuade, w should educate or inform people about choices or options tha may make a difference in their health and the quality of thei lives. It is important to provide perspective, to provide informa tion that will alert or alarm those at highest risk and reassur those at lowest risk, as health communicator Peter Sandman ha said. I like the idea of telling people their chance of not gettin cancer, rather than focusing solely on the risk of getting cancer There needs to be balance in the messages that one provides i the mass media as well as in tailored communications to reduc the noise volume and produce calmer, more reasoned cancer ris information.

Introduction of Section: Breakout Session Reports

J. Paul Van Nevel

The Cancer Risk Communication workshop was designed to share knowledge and expertise across professional disciplines, to identify effective methods of communicating risk information, and to establish a research agenda for cancer risk communications. There were two primary workshop objectives:

- Identify "best practices" in cancer risk communications for immediate use by risk researchers, journal editors, institutional communicators, and mass media representatives, as well as those who convey risk information through interpersonal communication.
- Identify the most pressing research needs in cancer risk communications.

After hearing 2 days of presentations, workshop participants divided into two working groups—one to "best practices" and the other to identify research needs for cancer risk communications. Following are the reports of the two working groups. Elaine Bratic Arkin summarized the results of the working group on best practices, and Edward Maibach, Ph.D., summarized the results of the group that considered a research agenda.

Workshop organizers hope that ideas in these two papers can be implemented rapidly. Many of the best practices can be implemented immediately. The research needs can be addressed not only by organizations that fund such research but also by investigators from a variety of professional disciplines.

Affiliation of author: Office of Cancer Communications, National Cancer Institute, Bethesda, MD.

Correspondence to: J. P. Van Nevel, National Institutes of Health, Bldg. 31, Rm. 10A31, MSC 2580, Bethesda, MD 20892 (e-mail: nevelp@occ.nci.nih.gov).

Cancer Risk Communication—What We Need to Learn

Edward Maibach

Participants were charged with "identifying the most pressing and important basic and applied research questions in cancer risk communication." Specifically, participants were asked to address two questions: 1) Which underlying processes in cancer risk communication need to be better understood to advance the field? and 2) How can we advance our ability to conduct theory-based risk communication research in the real-world context of cancer information? A third question was advanced in the course of the discussion: What research is needed to help accelerate the adoption of "best practices" in primary care and other settings in which cancer risk information is routinely communicated?

Underlying Processes

The Roles of Cognition

Although there is a considerable amount of literature on cognitive aspects of risk perception and communication, we a need to enhance our understanding of and ability to operationalize these cognitive functions and to better represent their multidimensional nature.

1) How do perceptions of risk vary among individuals at the same, and at different, levels of actual risk?
2) Which dimensions of risk most influence cognitive, affective, and behavioral responses to risk information, and how should they be measured?
3) How do semantic labels affect perceptions of risk and subsequent behavior (e.g., mammograms can be labeled as a way to detect breast cancer early or as a way to routinely affirm healthy breasts)?
4) How do cognitive constructs, such as mortality salience, illness representations, and perceived illness time frames, influence perceptions of, and responses to, cancer risk?

The Roles of Affect

Current theories of risk perception and communication tend to be highly cognitive, yet affective states moderate how risk messages are processed, are a consequence of risk information processing, and mediate subsequent cognitive and behavioral responses to risk information. A better understanding of the affective components of risk perception, risk response, and risk communication, and their relationship to cognitive components, are needed.

1) How can theories of risk perception and communication be modified to better account for the roles of affect?
2) What are the cognitive and emotional consequences to an individual faced with a risk for which there is no known intervention (e.g., a genetic risk)?
3) How can defensive or avoidance reactions be minimized?

The Roles of Values

Values play an important but poorly understood role in many risk communication situations. If one potential objective of a risk communication program is to help people make good in-

Correspondence to: Edward Maibach, Ph.D., M.P.H., Porter Novelli, 1120 Connecticut Ave., N.W., Suite. 1100, Washington, DC 20036 (e-mail: emaibach@porternovelli.com).

© Oxford University Press

formed decisions on the basis of their values, the question arises: How do we help people identify those values? Moreover, people's health "values" (i.e., the relative priority assigned to preservation of health versus other possible priorities) have been known to change in response to serious health threats. A better understanding of the mechanisms of this change may be important to better understanding the structure of effective cancer risk communication.

1) How can individual decision makers be aided in assessing and employing their own values to make the best possible decision for their current situation?
2) What are the experiences that precipitate a change in health values?

Information Searching

A better understanding of individuals' information search processes is critical to an enhanced understanding of the attributes of information that are perceived to bc relevant. An enhanced understanding of this domain is likely to have implications both for theories of risk perception and response, as well as for designing more effective risk communications.

1) What are the processes underlying how people search for risk information? How do their biases influence the process and outcome? What types of information do they seek (e.g., do they seek information that confirms previously held notions or information that may lead them in new directions)?
2) How does information gathered during the search affect subsequent experiences related to quality of life (e.g., information about the implications of a positive genetic test)?

Developmental Differences

Much, if not most, of what we know about risk perception and the relationship between risk perceptions and subsequent actions comes from the study of adults. Developmental differences undoubtedly influence the way in which children and adolescents process and act on risk information. Systematic examination of these differences is required to better understand the unique challenges of more effectively communicating with young audiences.

1) How do developmental differences influence cognitive and affective processing of health risk information?
2) What are the appropriate outcomes of risk communication for youth audiences?

Building Risk Communication Research Capacity

Infrastructure

A number of strategies were identified to enhanee the research infrastructure and thereby accelerate the pace and quality of cancer risk communication research. Successful examples of each of these opportunities are seen in other fields of research.

Development of a population-based information delivery "test bed" that would be available to different research teams would encourage and facilitate theoretical and applied health risk communication research. An interactive kiosk system developed by Strecher and colleagues *(1)*, installed in 100 public spaces in Michigan, was used as an example of an information delivery system that could be used as a shared resource amon many investigators.

Development of a nationally representative survey, con ducted on a recurring basis with input from investigators in th field and with data that would be publicly available, would pro vide a natural vehicle through which to assess the relationshi among risk perceptions, antecedents, and consequences ove time. The survey, similar to the General Social Survey or th National Health Interview Survey, would provide a tremendou resource for the development and testing of risk communicatio theory.

New communication technologies are providing dramaticall expanded opportunities for conducting risk communication re search, yet most professionals with an interest in health ris communication have limited access to training that will allov them to take advantage of these developments. Funding existin laboratories for health communication research to conduct short format or longer-format training programs for research profes sionals would increase the number of research teams capable o taking advantage of these research opportunities.

Cancer risk communication research is inherently an inter disciplinary activity, yet few extant research teams are ideall staffed across the various relevant disciplines. Research fundin mechanisms should be used to encourage innovative interdisci plinary staffing, wherever possible.

Mechanisms for tracking developments in the many academi disciplines from which cancer risk communication draws ar seriously underdeveloped. A great need exists to use electroni publishing to broadly disseminate abstracts and other relevan information among research professionals working on healt risk communication issues.

Outcome Evaluation

With the broad, multifaceted nature of risk communication more consideration must be given to the potential objectives o risk communication and how accomplishment of those objec tives can best be assessed in response to a risk communicatio program.

1) What is the desired outcome of risk communication and wha criteria are we going to use to demonstrate that a risk com munication is effective? What is the appropriate role for im pact on sense of well-being and quality of life?
2) How do the objectives of the risk communication effort in fluence the evaluation design? There is a need to move be yond using specific behavioral outcomes as the evaluatio end point, especially in situations in which the communica tion objective is to enable informed decision making by en couraging individuals to consider potential risks and benefit in the context of their own values.

Research to Accelerate Identification and Adoption of Best Practices

Primary Care Setting

For a variety of reasons, many people report that a health car provider, often a primary care provider, is their preferred source of important health information, yet the literature also indicates that health care providers, and the primary care setting, often fai to deliver effective health risk information. A better understand ing of these barriers and the means for overcoming them ar

required to make better use of one of the preferred channels of health risk information by consumers.

1) How are primary care providers currently providing cancer risk information?
2) What constitutes successful cancer risk communication in a primary care environment? What does the process look like, and what are the desired outcomes?
3) What are the barriers in primary care to successful provision of cancer risk information? Which of those barriers can be modified and how?
4) Realistically, how can the training of primary care providers be modified to better enable them to addresses their patient's individual information and psychosocial needs?

Facilitating Decision Making

Fischhoff *(2)* made the case, and participants largely concurred, that in risk communication less is more. An act of risk communication is more likely to be effective to the extent that it presents a parsimonious amount of information—new or otherwise—that best allows members of the audience to understand the risk as an expert would. This perspective poses a number of important research questions.

1) What are efficient and effective means for identifying the most relevant information to present?
2) What kinds of decision-supporting interventions (e.g., decision trees or aids) are most effective with which types of decisions and individuals?
3) What mechanisms are most effective in clarifying an individual's information needs in advance of partaking in the information search process (e.g., to prepare for a discussion with a physician)?
4) What does the public see as the most important health decisions with which they are faced? What do they see as the barriers and facilitators they face in making what they consider to be a good decision? What type of support do they need to improve their decision making and to believe that they made a good decision?

Effective Channels

The range of channels that can be used to communicate risk information is large and growing. As the options expand, it becomes increasingly important to understand the strengths and limitations of each of these channels.

1) What is the most effective means for presenting risk information in each of the various channels currently available?
2) Which channels are particularly well suited for which types of risk communication?
3) Given that motivated members of the lay public often seek out publications intended for health professionals, will specifically crafted lay summaries of the information enhance outcomes at the individual and the population level?
4) How do we harness social diffusion of important risk information?
5) What are the most effective means for enhancing exposure among affected audiences to important health risk information?

Message Content

Ultimately, risk communicators must make difficult decisions about what information is most important to present. New information technologies are providing dramatically enhanced options to present different information to different types of people. Even the new technologies have limits, however, and a better understanding of the most relevant message-tailoring variables is important to advance the field. Moreover, risk information is inherently uncertain. Determining how best to communicate this uncertainty without undermining the effectiveness of the message is an important outstanding question.

1) Which individual differences are most critical to consider in creating tailored risk communication messages (e.g., level of risk exposure)?
2) What are the most appropriate ways to communicate risk uncertainty to the public?

Conclusion

The extant risk communication literature has clear relevance and value for better understanding the challenges of cancer risk communication. The nature and scope of the research questions identified, however, demonstrate that there is a pressing need for additional research so that the benefits of the expanding knowledge base on cancer risk can be more broadly and more effectively employed by members of the population to make personally relevant health decisions. Understanding the scope of these research questions will help investigators and funding agencies understand where the legitimate future research opportunities lie.

References

(1) Strecher VJ, Greenwood T, Wang C, Dumont D. Interactive media and risk communication. Monogr Natl Cancer Inst 1999;25:134–9.

(2) Fischhoff B. Why risk communication can be so difficult. Monogr Natl Cancer Inst 1999;25:7–13.

Cancer Risk Communication—What We Know

Elaine B. Arkin

Introduction

Participants in this session were charged with identifying "best practices" in cancer risk communication for mass media and interpersonal communication. The participants discussed a variety of factors that contribute to effectively communicating about cancer risks, beginning with understanding the message receiver. Participants explored the continuum of steps and decisions that contribute to best practice and how to ensure that decisions made in each stage of this complex communication process contribute to effective communication.

Background

The public is fed a steady diet of risk information. For cancer, this includes information about the risks of certain cancers (e.g., breast cancer), lifestyles that increase the odds of getting cancer (tobacco use), cancer-related environmental exposures (asbestos), risk-reduction practices (eating five fruits and vegetables a day), and risk trade-offs (medical treatments that may increase cancer risks).

Information about risks is communicated in a variety of ways: Research findings are reported in scientific journals and then in the news; public policy deliberations are shared through conferences and published reports; and planned communication campaigns address longer-term problems, such as asbestos exposure and radiation fallout.

For the mass media, news about risks can be an attention getter, so it is an attractive kind of story to cover. Often, the risk news story is incomplete—partially because of inherent time and space limitations and partially because the complexities of risks could not practically be explored in most news venues. Even explaining these complexities to news reporters is a challenge; for example, the National Cancer Institute receives over 12 000 media inquiries each year, almost always with short response deadlines that limit the depth of information that can be provided or used. In addition, the number of channels that provide access to news and health information is increasing rapidly, with an explosion of cable television programs and World Wide Web sites offering health information.

As a result, the public may be confused, unduly alarmed, or too jaded to respond appropriately to serious risks. When people underestimate or overestimate their risk, not only individual decisions but also public policy decisions can be adversely affected. Therefore, participants identified a need to construct risk communications that convey adequate information about a health risk and to help ensure that the public is enabled to respond in appropriate ways.

Complexities of Risk

In general, the term "risk communication" is used to cover a broad range of types of risks and types of communication. The origin of the health risk may be environmental, genetic, or behavioral, or it may be related to medical treatment. Each of these sources of risks may engender different degrees of understanding and emotions among the public.

Often, a risk results from multiple causes, further complicating both communications and comprehension. How the public receives communications about risks having multiple causes has not been well documented. And, although the public's understanding of and reaction to other types of risks (e.g., environmental and behavioral) have been explored in depth, public response to cancer risks has been less well documented and is complicated by the public's cancer fears.

Where there have been similar situations in the past, the risk communicator can anticipate the potential for public reaction, such as appropriate or inappropriate alarm, action, and demand for risk reduction (e.g., the public may be unduly alarmed in response to news about a potential but rare cancer risk in processed food). Although it is not possible to foresee specifically how the public will react, understanding how the public has responded to similar events in the past is helpful in planning cancer risk communications. In addition, past experience can guide formative research conducted to improve the communication or anticipate the public's response.

Conversely, new risk information related to entrenched personal behaviors may be ignored, denied, or discounted. In such cases, the risk communication planner may need to consider broader, longer-term communication programs with the use of multiple strategies to impress affected individuals.

Therefore, participants noted, important first steps in effective risk communication include identifying how the public has reacted to similar types of risk information in the past, conducting market research to better anticipate public response to the new risk (where time and resources permit), and determining appropriate communication strategies, such as providing background information to reporters, providing the interested or alarmed public with access to accurate information as well as advice about steps to take if they are worried about risk exposure, or planning a longer-term program to influence public attitudes or behaviors.

Suggestions for Improving Risk Communication Practice

Communicating effectively about cancer risks requires interdisciplinary best practice, spanning communications by risk and health communication practitioners, health care providers, public health officials, journal editors, public relations practitioners, reporters, advocacy groups, and scientists.

Even though there still are many uncertainties regarding certain aspects of how to effectively communicate about risks, the group concluded that best practice calls for tracing the path of communication from the researcher to the intended recipient and for making a series of judgments related to the purpose, context, and audience for the communication.

In the absence of planning models specific to risk communications, health communication planning models (such as the

Correspondence to: Elaine B. Arkin, Porter Novelli, 3435 N. 14th St., Arlington, VA 22201.

© Oxford University Press

health communications "wheel" developed by the National Cancer Institute) should be used, participants said *(1,2)*.

To increase effective risk communication, participants specifically suggested the following:

- Establish a "Center of Excellence" in risk communication (i.e., at the National Cancer Institute) to serve as a source for research, training, technical assistance, and guidelines setting.
- Promulgate principles of best practice on the basis of standards in related fields, including health education and health communications, and tailor them for different types of risk communicators (including journalists, journal editors, institutional spokespersons, and researchers).
- Provide a "guide to communicating about cancer risks" for health care providers faced with patient counseling responsibilities (such as counseling about genetic testing).
- Promote the use of "checklist" tools to guide immediate decisions (such as writing press releases or answering reporter and public inquiries) when brevity in response and response time is the rule.
- Prepare guidelines on how best to present risks in numerical terms to the public for use by scientists preparing or reviewing journal articles and for journal editors, reporters, and others interpreting risks for the public.
- Create primers, checklists, or other tools to help the public understand risks (participants cited guides developed by organizations, including the National Breast Cancer Coalition).
- Work with professional organizations, such as the National Association of Science Writers, to develop a consensus about necessary components of a risk press release or feature article as a benchmark.
- Work with scientific journals to encourage authors to prepare a "public abstract" or context (such as prior studies and baseline data), covering reported work that would help explain the meaning of results to the public (including, participants noted, professionals from other disciplines).
- Work with scientific journals to construct and promote consistent ways of presenting data and encourage journals to require authors to meet certain standards.
- Work with radio or other mass media to produce and air regular short features covering and explaining reports in medical and scientific journals.
- Provide additional "contextual documents" to intermediaries, such as journal editors who publish the results of studies (these documents should provide background or context for the new findings).
- Cross-train between journalists and scientists to help improve understanding of how each communicates about risks and to improve interaction to meet each other's needs.
- Provide the suggested interaction, training, and guidance for risk communication within the broader framework of accepted health communication practice (such as the need for formative research and evaluation).

The group also identified the following gaps in guidance for effective risk communication practice:

- No planning framework that identifies key decisions in risk communication planning and evaluation
- No body of evaluation research that provides best praetice examples
- Little understanding of differences in comprehension and decision making related to cancer risks among different cultural groups
- Little knowledge of how the public perceives and processes information about multiple risk factors

Guidelines for Best Practice

What can be considered best practice varies by type of risk communication. For example, some best practices among reporters differ from those of institutional public relations managers. And the most effective ways of communicating about fast-breaking news differ from the ways to plan effective communication campaigns for risk reduction. Nevertheless, participant discussion revealed the following general guidelines to effective communications about cancer risks:

1) First identify the risk communication goal.
2) Determine who needs to know.
3) Factor in target audiences' perceptions of risk.
4) Select the most appropriate channels for each risk situation.
5) Determine how to manage multiple voices.
6) Explain the risk clearly.
7) Provide context to improve understanding.
8) Include recommended actions or other advice.
9) Assess the outcome.
10) Determine whether the situation requires additional attention.

These guidelines are particularly pertinent for planning a proactive educational campaign, and some are useful when dealing with episodic reports of risk in the media.

First Identify the Risk Communication Goal

Identifying the goal (purpose) of the risk communication is an important first step. Answering the goal-setting question "Why do we want to communicate about this risk?" should guide subsequent decisions about how the risk will be communicated. Diverse communicators often have different goals for reaching out to the public about a cancer risk; some may want to minimize the reaction to a risk, whereas others may want to maximize it. Examples of goals participants cited include the following:

- Inform the public about a risk (the public's right-to-know).
- Help people more accurately understand their own risks.
- Increase individuals' abilities to make informed decisions about reducing risks.
- Decrease public misconceptions about certain risks and provide reassurance when appropriate.
- Inform public policy decisions.
- Persuade the public to support legislation, public policy, or resource allocation to help remedy a problem.
- Promote the capabilities, activities, or policies of a corporation or other entity.
- Increase support for risk-related research.

Determine Who Needs to Know

Participants were quick to point out that there is no uniform "general public," but rather numerous smaller, sometimes intersecting groups of individuals sharing common concerns, interests, perspectives, or demographic characteristics. For example, individuals can be clustered into narrower "target audiences" according to degree of risk exposure or ability to influence risk reduction policies. Decisions about "who needs to know" should take into account not only who is affected by the risk, but also

who can affect the risk situation or who is not affected but unduly concerned. Individuals who interact with those at risk (such as caregivers or family members) are one example. Others include policymakers, public health officials, health care pro viders, and business managers who may be called into the situation.

For each group identified, there will likely be some differences in what needs to be communicated. Also, each group may bring different levels of risk understanding and different perspectives and emotions that need to be considered in risk communication design.

Factor in Target Audiences' Perceptions of Risks

As addressed in the panel on "Overarching Considerations in Risk Communication" [as well as by other presenters *(3)*], an individual's own perceptions of his or her risks and life situation influence how risk information is received and used.

First, the public's fear of cancer needs to be acknowledged. Fears can lead to denial, avoidance of appropriate responses, or overreactions to new risk information. Individuals' perceptions of the personal relevance of the risk as well as their perceptions of their own ability to protect themselves can affect how they receive and respond to the risk communication. Previously held misconceptions about cancer and cancer risks also can color reactions to new risk information.

The public frequently demands answers to specific risk questions. Therefore, they should not be viewed solely as the end target of risk communications, said participants. Rather, public perceptions, wants, and needs should be identified early on and integrated into risk communication planning. Therefore, as one participant noted, "[Best practice] is not a top-down approach . . . but it's formative research [with] our target audience."

Select the Most Appropriate Channels for Each Risk Situation

Best practice includes weighing channel options and selecting the most appropriate channels for each particular risk situation. Where the risk message appears can affect how the public receives it. Participants noted that, whereas the generation of new knowledge occurs within the scientific community, the translation of the new information to the public entails many steps, each with inherent communication challenges. Information provided through one-on-one counseling by a health care provider to a patient will be perceived differently from news in the mass media, for example. In most cases, information delivered in person by a health care provider is more credible and influential than the evening news.

The risk communicator should recognize the attributes and limitations of different channels, including the varying degrees of credibility they lend to risk messages. Usually, participants noted, risk communication is conveyed through multiple types of channels, whether intentionally (planned) or unintentionally.

Once channels are identified, steps should be taken to work with the "gatekeepers" to those channels (such as news reporters, journal editors, or health care providers).

Determine How to Manage Multiple Voices

Another complicating factor is the variety of spokespersons potentially involved in communicating about risk. Participants provided examples, including health care providers, scientists, corporate spokespersons, advocates, and public relations practitioners (such as those at medical institutions).

In addition to those who personally communicate about risks, there also are other decision makers who influence how risk information is communicated, such as journal editors, news directors, senior scientists, and corporate directors. Each may have a different perspective on the information being provided, different levels of preparedness to communicate complex information to the lay person, and different degrees of credibility with the public. As one participant noted, "The credible source is the most important factor of communication."

Several participants mentioned a dilemma in many organizations, especially those related to environmental policies and research, in which the communication specialist (the individual most likely to be skilled in communicating about risks) is "marginalized"—i.e., not seen as a "central player" in planning how to communicate about cancer risks.

It also was noted that there are inherent tensions in discussing the goals of risk communications across disciplines; for example, public health professionals may look at emerging evidence of a threat to health and determine a need to take action to safeguard the public's health, whereas the scientific community may perceive a need to wait for a greater level of scientific evidence or proof of risk before taking a stand. Another dimension is added when there are differences in perspectives (or vocal opposition) between public interest groups, the scientific community, and business interests, as in the case of alar contamination of apples.

This means that communication about any specific risk is likely to occur within an environment replete with confusion—including varied, and sometimes conflicting, information and advice.

Therefore, the risk communication practitioner needs to identify who the communicators will be; determine differences in perspectives, goals, and potential to affect public risk response; and determine how best to manage the potential effect of "multiple voices" on the outcome of the risk communication.

Explain the Risk Clearly

Participants discussed several complicating factors related to crafting risk communication messages.

- Best practice requires synthesizing aspects of many disciplines that contribute to effective risk communieation (ineluding health communications, behavioral sciences, risk management, political science, journalism, social marketing, medicine, epidemiology, and psychology).
- Many individuals with the responsibility are trained in only one of these disciplines, leading to variations in how risk communications are shaped.
- Risk concepts are often complex and difficult to explain.

The construction of risk communication messages has been explored in depth, although such exploration does not necessarily produce generalizable guidance for all risk communications, given the variation among risk situations. Participants discussed the attributes of tailored messages, guidelines for journalists for the reporting of cancer risks, and tips for writing and presenting cancer risk information to improve comprehension and to inform decision making. (*See* papers in this monograph by Rimer and Glassman, Russell, Brody, Schwartz and Woloshin, Lipkus and Hollands, O'Connor et al., and Strecher et al.)

For example, the National Cancer Institute *(4)* recently conducted a series of focus groups to explore how terminology selected (e.g., substituting the word "chance" for "risk" under

some circumstances) and how presenting risk information visually can help improve public understanding.

As attributes of best practice, participants discussed the need to ensure that risk information is both comprehensible and is provided in context (e.g., with information that helps explain the significance of the risk for an individual or community), whether the risk can be avoided, as well as the consequences of risk exposure (*see* Table 1).

Provide Context to Improve Understanding

The public is more likely to understand and respond appropriately to risk information when "background" information explains who is at risk and who is not, how significant this risk is in comparison with other health threats, how dire the consequences are, and how sure scientists are that the risk can lead to cancer *(5,6)*. This type of context can help people decide whether and how the risk is personally relevant.

Explaining uncertainties is important, participants said, because, as scientific evidence continues to accumulate, judgments about risks and disease risk factors change. If the public does not understand that these changes are a positive outcome of the scientific process, the credibility associated with risk communications can be undermined.

Several participants also mentioned the importance of explaining limitations associated with a specific risk communication, such as reporting the outcomes of one study as opposed to a cumulative body of data.

The type of contextual information to be provided should be, in part, influenced by the public's current views on the topic; e.g., where it is known that the public has little knowledge of the risk, more definitional information may be required than when referring to a risk about which the public is familiar (such as the risk of lung cancer associated with smoking).

Include Recommended Actions or Other Advice

Informing the public without explaining what can be done about a cancer risk can lead to frustration or even anger. When possible, risk communications should include the following:

- Recommended actions to avoid the risk or lessen the consequences of exposure
- Sources of additional information or assistance

Participants discussed the benefits of using decision aids to help individuals process information, clarify options, and make appropriate decisions to ameliorate the effects of risk exposure.

Table 1. Crafting risk messages

- Choose simple, nontechnical terms; define unavoidable terminology.
- Explain the risk in absolute terms where possible; avoid using only relative risk comparisons, which are confusing to some people.
- Recognize that multiple and cumulative risks are difficult for many people to understand; if there are multiple risks, explain each in the same way.
- Try to explain the dimensions of the risk in several ways to account for differences in how people process information (e.g., numerically, visually, and explanatory text).
- Provide context (e.g., findings of previous studies).
- Explain how serious the risk situation is, including the consequences of risk exposure.
- Define who is at risk.
- Include implications and any recommended action steps.
- Include information about how to reduce risks.
- Point out any uncertainties or gaps in information to avoid distrust.
- For research results, include who funded, who conducted, and where the study findings were published.

Assess the Outcome

People do not respond to risk information in a consistent manner. Therefore, it is important to assess how the risk communication was received and how the public reacted. In this case, participants said, "the public" includes both target audiences and others who were exposed to the risk information.

Assessment should indicate the degree to which the original purpose or goal has been met. Assessment also can reveal unintentional results, such as individuals becoming unduly alarmed or taking unnecessary precautions, or otherwise misinterpreting or misusing the risk information.

Assessment is needed to make decisions about whether additional communications (or other actions) are needed to address the risk situation.

Determine Whether the Situation Requires Additional Attention

Assessment information should be used to decide whether the original goals have been met. Determinations also can be made about whether to continue or change tactics to address changes in public perceptions, actions, or risk exposure.

Risk situations change over time, and the risk communication approach should be altered to fit changed circumstances; e.g., further study can change what is known about a risk, or new policies and programs can change how the risk affects the public. Political forces or advocacy groups may have established or changed positions. Also, public knowledge and risk-related actions are likely to shift as a result of risk communications. Therefore, periodic reassessment of a risk situation should be conducted to determine whether, why, how, and what to communicate about a risk.

Summary

In addition to these broad guidelines, participants suggested that more specific guidance be provided for risk communicators in different sectors, including journalists, journal editors, governmental and organizational spokespersons, and risk researchers. The National Cancer Institute and other organizations also have produced materials to help the public understand, interpret, and make decisions about new risk information and to guide those who communicate about cancer risks. These kinds of efforts to enable the public to apply risk information are vital and should be continued, participants said.

References

(1) National Cancer Institute. Making health communication programs work: a planner's guide. Bethesda (MD): National Cancer Institute; 1992.

(2) Centers for Disease Control and Prevention. CDCynergy, Version 1 [CD-ROM]. Atlanta (GA): Centers for Disease Control and Prevention; 1998.

(3) Fischhoff B. Why (cancer) risk communication can be hard. J Natl Inst Monogr 1999;25:7–13.

(4) National Cancer Institute, Office of Cancer Communications. How the public perceives, processes, and interprets risk information: findings from focus group research with the general public. Bethesda (MD): National Institutes of Health, National Cancer Institute; 1998 Report No.: POS-T086.

(5) Rothman AJ, Kiviniemi MT. Treating people with information: an analysis and review of approaches to communicating health risk information. J Natl Cancer Inst Monogr 1999;25:44–51.

(6) Russell C. Living can be hazardous to your health: how the news media cover cancer risks. J Natl Cancer Inst Monogr 1999;25:167–70.

NIH LIBRARY
3 1496 00816 8554

Matching Strength of Message to Strength of Evidence: a Discussion 85
Barnett S. Kramer

Introduction of Section: 88
Persuasion for the Purpose of Cancer Risk Reduction: Understanding Responses to Risk Communications
Lisa G. Aspinwall

The Effect of Risk Communication on Risk Perceptions: the Significance of Individual Differences 94
Meg Gerrard, Frederick X. Gibbons, Monica Reis-Bergan

Risk Perception and Risk Communication for Cancer Screening Behaviors: a Review 101
Sally W. Vernon

Persuasion for the Purpose of Cancer Risk Reduction: a Discussion 119
Peter Salovey, Tamera R. Schneider, Anne Marie Apanovitch

Introduction of Section: 123
Implications for Improving Risk Communication Through Various Channels
J. Paul Van Nevel

Risk Communication in Clinical Practice: Putting Cancer in Context 124
Lisa M. Schwartz, Steven Woloshin, H. Gilbert Welch

Interactive Multimedia and Risk Communication 134
Victor J. Strecher, Todd Greenwood, Catharine Wang, Dana Dumont

Is There a Use for Tailored Print Communications in Cancer Risk Communication? 140
Barbara K. Rimer, Bernard Glassman

The Visual Communication of Risk 149
Isaac M. Lipkus, J. G. Hollands

Risky Business—Communicating Scientific Findings to the Public 163
Jeffrey Levine

Living Can Be Hazardous to Your Health: How the News Media Cover Cancer Risks 167
Cristine Russell

Communicating Cancer Risk in Print Journalism 170
Jane E. Brody

Challenges to Improving Health Risk Communication in the 21st Century: a Discussion 173
Geoffrey T. Fong, Lynn A. Rempel, Peter A. Hall

Implications for Improving Risk Communication Through Various Channels: a Discussion 177
Cristine Russell

Introduction of Section: 179
Breakout Session Reports
J. Paul Van Nevel

Cancer Risk Communication—What We Need to Learn 179
Edward Maibach

Cancer Risk Communication—What We Know 182
Elaine B. Arkin

MONOGRAPHS
JOURNAL OF THE NATIONAL CANCER INSTITUTE

Innovative Study Designs and Analytic Approaches to the Genetic Epidemiology of Cancer

1999
Number 26

Contents

Introduction
Daniela Seminara

I. GENE DISCOVERY PANEL

Discovery of Cancer Susceptibility Genes: Study Designs, Analytic Approaches, and Trends in Technology 1
Daniel J. Schaid, Ken Buetow, Daniel E. Weeks, Ellen Wijsman, Sun-Wei Guo, Jurg Ott, Carol Dahl

II. GENE CHARACTERIZATION PANEL

Design of Gene Characterization Studies: an Overview 17
Duncan C. Thomas

Case–Control Studies of Common Alleles and Environmental Factors 25
Neil Caporaso, Nathaniel Rothman, Sholom Wacholder

Family-Based Association Studies 31
W. James Gauderman, John S. Witte, Duncan C. Thomas

Cohort Studies for Characterizing Measured Genes 39
Bryan Langholz, Nathaniel Rothman, Sholom Wacholder, Duncan C. Thomas

Multistage Sampling for Disease Family Registries 43
Kimberly D. Siegmund, Alice S. Whittemore, Duncan C. Thomas

Detection of Interaction Involving Identified Genes: Available Study Designs 49
Alisa M. Goldstein, Nadine Andrieu

Kin-Cohort Designs for Gene Characterization 55
Mitchell H. Gail, David Pee, Raymond Carroll

III. INTEGRATION PANEL

Study Design in Genetic Epidemiology: Theoretical and Practical Considerations 61
Alice S. Whittemore, Lorene M. Nelson

Integrated Designs for Gene Discovery and Characterization 71
Lue Ping Zhao, Corinne Aragaki, Li Hsu, John Potter, Robert Elston, Kathleen E. Malone, Janet R. Daling, Ross Prentice

IV. APPLICATIONS PANEL

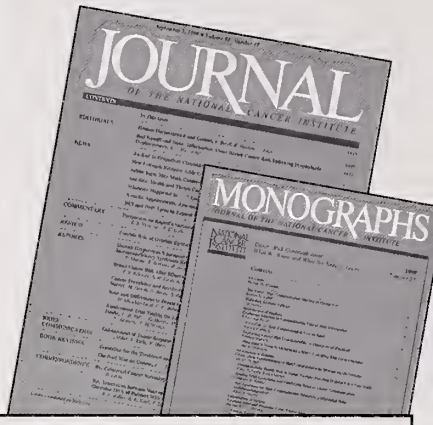
Confirmation of Prostate Cancer Susceptibility Genes Using High-Risk Families 81
Gail P. Jarvik, Janet L. Stanford, Ellen L. Goode, Richard McIndoe, Suzanne Kolb, Mark Gibbs, Leroy Hood, Elaine A. Ostrander

Study-Design Issues in the Development of the University of Southern California Consortium's Colorectal Cancer Family Registry 89
Robert W. Haile, Kimberly D. Siegmund, W. James Gauderman, Duncan C. Thomas

Design and Analysis Issues in a Population-Based, Case–Control–Family Study of the Genetic Epidemiology of Breast Cancer and the Co-operative Family Registry for Breast Cancer Studies (CFRBCS) 95
John L. Hopper, Georgia Chenevix-Trench, Damien J. Jolley, Gillian S. Dite, Mark A. Jenkins, Deon J. Venter, Margaret R. E. McCredie, Graham G. Giles

Combined Influence of Genetic and Dietary Factors on Colorectal Cancer Incidence in Japanese Americans 101
Loïc Le Marchand

Subscribe to the leading journal in primary oncology research!



The *Journal of the National Cancer Institute* publishes the best papers in oncology from around the world.

Internationally acclaimed as the source for the most up-to-date information from the rapidly changing fields of cancer research and treatment, the *Journal of the National Cancer Institute* is available by subscription from Oxford University Press.

Each issue is packed with news about important cancer-related science, policy, politics, and people. And you'll also find editorials and reviews by experts in the field, book reviews, and commentary on timely topics.

Online Version

www.jnci.oupjournals.org

And there's more!

Immediately on subscribing, you will become a member of the NCI Information Associates Program, which includes these outstanding benefits:

- One copy of each new *Journal of the National Cancer Institute Monograph**
- 24-hour access to PDQ®, the NCI's comprehensive cancer information database
- Listserv Subscription Service
- *ProtoCall*™ Newsletter
- Member Help Line Services

NATIONAL CANCER INSTITUTE
INFORMATION ASSOCIATES PROGRAM

For more information about the Information Associates Program, call 1-800-624-7890 (+301-496-7600 outside the USA).

*Free to members on request prior to publication.

I wish to subscribe to the *Journal of the National Cancer Institute.*

Name (please print)_____

Address _____

_____

ZIP/Post code_____ Country_____

❒ I enclose the correct amount (check payable to Oxford University Press).

❒ Please bill me.

Please charge my

❒ MasterCard ❒ Visa ❒ American Express ❒ Diners ❒ JCB card.

Card number

Exp. date_____ Signature _____

2000 Subscription Rates Volume 92 (24 issues)

† institutional subscribers will now receive all monographs automatically as they are published with no need to order separately. If, however, you would prefer not to receive this automatic service, please contact your nearest Journal Subscriptions Department at Oxford University Press. Personal subscribers will have the opportunity to order monographs free of charge prior to publication.

Individuals	
US & Canada	❒ One year: US$155
UK & Europe	❒ One year: £150
Elsewhere	❒ One year: US$215
Institutions	
US & Canada	❒ One year: US$235†
UK & Europe	❒ One year: £210†
Elsewhere	❒ One year: US$265†

Single issues are available on request. Air-expedited delivery fee is included for subscriptions outside North America. Individual subscriptions must be paid for by personal check or credit card and sent to private addresses only. Canadian residents add 7% GST.

To begin your subscription, call 1-800-852-7323 (USA & Canada). Or call 919-677-0977. Fax: 919-677-1714.
Outside the Americas, call +44 (0)1865 267907 or fax +44 (0)1865 267485.

In the Americas, please return to:
Oxford University Press, Journals Marketing (jnl), 2001 Evans Road, Cary, NC 27513, USA
Elsewhere, please return to:
Oxford University Press, Journals Marketing (jnl), Great Clarendon St., Oxford, OX2 6DP, UK

OXFORD UNIVERSITY PRESS

MONOGRAPHS

JOURNAL OF THE NATIONAL CANCER INSTITUTE

Number 26 *1999*
ISSN 0027-8874
ISBN 0-19-922480-3

LIBRARY
JAN 03 2000
National Institutes of Health

Barnett S. Kramer
Editor-in-Chief

J. Gordon McVie
European Editor

Eric J. Seifter
Book Reviews Editor

J. Paul Van Nevel
News Editor

Frederic J. Kaye
Douglas L. Weed
Reviews Editors

Martin L. Brown
Economics Editor

ASSOCIATE EDITORS

Susan G. Arbuck
Frank M. Balis
William J. Blot
Peter M. Blumberg
John D. Boice, Jr.
Louise A. Brinton
Ross C. Donehower
Susan S. Ellenberg
Ellen G. Feigal
Suzanne W. Fletcher
Michael A. Friedman
Patricia A. Ganz
Edward L. Giovannucci
John K. Gohagan
Frank J. Gonzalez
Michael M. Gottesman
Peter Greenwald
Jean L. Grem
Donald E. Henson
Susan M. Hubbard
Colin R. Jefcoate
Arnold D. Kaluzny
Frederic J. Kaye
Hynda K. Kleinman
Theodore S. Lawrence
Bernard Levin
W. Marston Linehan
Marc E. Lippman
Scott M. Lippman
Darrell T. Liu
Dan L. Longo
Reuben Lotan
Douglas R. Lowy
Susan G. Nayfield
David L. Nelson
David G. Poplack
Ross L. Prentice
Alan S. Rabson
Edward A. Sausville
Robert H. Shoemaker
David Sidransky
Richard M. Simon
Michael B. Sporn
Maryalice Stetler-Stevenson
J. Paul Van Nevel
Douglas L. Weed
Noel S. Weiss

STATISTICAL EDITORS

Janet W. Andersen
Jacques Benichou
Donald A. Berry
Barry W. Brown
Timothy R. Church
Bernard F. Cole
Susan S. Ellenberg
Scott S. Emerson
Eric Feuer
Mary Foulkes
Mitchell H. Gail
Edmund A. Gehan
Barry I. Graubard
Sylvan B. Green
Susan G. Groshen
Li Hsu
Marek Kimmel
Victor Kipnis
James R. Murphy
Richard A. Olshen
Philip C. Prorok
Timothy R. Rebbeck
Philip S. Rosenberg
Harland N. Sather
Daniel J. Schaid
Mark R. Segal
Richard M. Simon
Donald M. Stablein
Robert E. Tarone
Sholom Wacholder

EDITORIAL ADVISORY BOARD

Hans-Olov Adami
Harry B. Burke
Pelayo Correa
Ethan Dmitrovsky
Merrill J. Egorin
Soldano Ferrone
Isaiah J. Fidler
Curtis C. Harris
Emily L. Harris
Kathy J. Helzlsouer
Maureen M. Henderson
Gloria H. Heppner
Allan Hildesheim
Waun Ki Hong
William J. Hoskins
Peter J. Houghton
David H. Johnson
V. Craig Jordan
Ilan R. Kirsch
Kenneth H. Kraemer
Alexandra M. Levine
Martha S. Linet
Henry T. Lynch
Frank L. Meyskens
Anthony B. Miller
Malcolm S. Mitchell
David L. Page
Kenneth J. Pienta
David Schottenfeld
Richard K. Severson
Jerry W. Shay
Debra T. Silverman
Sandra M. Swain
G. Marie Swanson
Mario Sznol
Raymond Taetle
Ian Tannock
Peter R. Twentyman
Regina G. Ziegler

Richard D. Klausner
Director, National Cancer Institute
Susan Molloy Hubbard
Director, International Cancer Information Center
Julianne Chappell
Chief, Scientific Publications Branch; Executive Editor

EDITORIAL STAFF

Scientific Editor: Daniela Seminara, Ph.D., M.P.H.

***Journal* Managing Editor:** Richard E. Manrow

Monograph Coordinator: Elizabeth Horowitz

Manuscript Editors: Elaine Price Beck
Janice Deal
Joan O'Brien Rodriguez
Elizabeth Horowitz

Editorial Assistants: Aylin An, Danielle D. Reed, Patricia A. Smith, Tenisha D. Smith

MARKETING STAFF

Marketing: Joy Cox

Press Contact: Dan Eckstein*

EDITORIAL POLICY: Manuscripts from key conferences dealing with cancer and closely related research fields, or a related group of papers on specific subjects of importance to cancer research, are considered for publication, with the understanding that they have not been published previously and are submitted exclusively to the *Journal of the National Cancer Institute Monographs.* All material submitted for consideration will be subject to review, when appropriate, by at least one outside reviewer and one member of the Editorial Board of the *Journal of the National Cancer Institute.* Opinions expressed by the authors are not necessarily those of the publisher or the editors.

Proposals for monographs should be submitted to the Editor-in-Chief, *Journal of the National Cancer Institute,* Oxford University Press, P.O. Box 31110, Bethesda, MD 20824-1110.

Journal of the National Cancer Institute Monographs are available on request to members of the National Cancer Institute Information Associates Program (one copy of each monograph per member). Monographs are also available from Oxford University Press, Journals Subscription Department, 2001 Evans Road, Cary, NC 27513 USA. Telephone toll-free in the U.S. and Canada, 1-800-852-7323, or 919-677-0977. Fax: 919-677-1714. E-mail: jnlorders@oup-usa.org.

© Oxford University Press

*Contact by telephone at 301-986-1891, ext. 112.

Innovative Study Designs and Analytic Approaches to the Genetic Epidemiology of Cancer

Sponsored by the National Cancer Institute
and the Fred Hutchinson Cancer Research Center

May 27–29, 1998
Seattle, Washington

List of Participants

Habibul Ahsan, M.D.
Columbia University
Nadine Andrieu, Ph.D.
Institut Gustave Roussy
Hoda Anton-Culver, Ph.D.
University of California
Corinne Aragaki, M.D.
Fred Hutchinson Cancer Research Center
Melissa Austin, Ph.D.
University of Wisconsin
Terri Beaty, Ph.D.
Johns Hopkins University
Norman Boyd, M.D.
Ontario Cancer Institute
Ken Buetow, Ph.D.
National Cancer Institute
Neil Caporaso, M.D.
National Cancer Institute
Chu Chen, Ph.D.
Fred Hutchinson Cancer Research Center
Georgia Chenevix-Trench, Ph.D.
Queensland Institute of Medical Research
Carol Dahl, Ph.D.
National Cancer Institute
Janet Daling, Ph.D.
Fred Hutchinson Cancer Research Center
Bob Day, M.D., Ph.D.
Fred Hutchinson Cancer Research Center
Mary Emond, Ph.D.
University of Washington
Ziding Feng
Fred Hutchinson Cancer Research Center
Pia Fish, M.P.H.
Fred Hutchinson Cancer Research Center
Andrew Freedman, Ph.D.
National Cancer Institute
Mitchell Gail, M.D., Ph.D.
National Cancer Institute
W. James Gauderman, Ph.D.
University of Southern California
Alisa Goldstein, Ph.D.
National Cancer Institute
John Grove, Ph.D.
University of Hawaii
Sun-Wei Guo
University of Minnesota
Robert Haile, Ph.D.
University of Southern California
Lee Hartwell, Ph.D.
Fred Hutchinson Cancer Research Center
John Hopper, Ph.D.
University of Melbourne
Li Hsu, Ph.D.
Fred Hutchinson Cancer Research Center
Gail Jarvik, M.D.
University of Washington
Damien Jolley, M.Sc.
University of Melbourne
Terri King, Ph.D.
University of Texas
Julia Knight, Ph.D.
Cancer Care Ontario
Loïc Le Marchand, M.D., Ph.D.
Cancer Research Center of Hawaii
Stephanie London, M.D., Dr.P.H.
National Institute of Environmental Health Sciences
Kathleen Malone, Ph.D.
Fred Hutchinson Cancer Research Center
Barbara McKnight, Ph.D.
University of Washington
Polly Newcomb, Ph.D.
Fred Hutchinson Cancer Research Center
Beth Newman, Ph.D.
University of North Carolina at Chapel Hill
Elaine Ostrander, Ph.D.
Fred Hutchinson Cancer Research Center

Jurg Ott, Ph.D.
Rockefeller University
Effie Petersdorf, M.D.
Fred Hutchinson Cancer Research Center
Ross Prentice, Ph.D.
Fred Hutchinson Cancer Research Center
Nathaniel Rothman, M.D., Ph.MHS
National Cancer Institute
Joseph Sambrook, Ph.D.
Peter MacCallum Cancer Institute
Daniel Schaid, Ph.D.
Mayo Clinic
Daniela Seminara, Ph.D., M.P.H.
National Cancer Institute
Linda Sharp, MSC
University of Aberdeen
Kim Siegmund, Ph.D.
University of Southern California
Janet Stanford, Ph.D.
Fred Hutchinson Cancer Research Center
Duncan Thomas, Ph.D.
University of Southern California
David Tritchler, Sc.D.
Ontario Cancer Institute
I-Ping Tu, Ph.D.
Stanford University
Thomas Vaughan, M.D.
Fred Hutchinson Cancer Research Center
Sholom Wacholder, Ph.D.
National Cancer Institute
Daniel Weeks, Ph.D.
University of Pittsburgh
Clarice Weinberg, Ph.D.
National Institute of Environmental Health Sciences
Alice Whittemore, Ph.D.
Stanford University
Ellen Wijsman, Ph.D.
University of Washington
John Witte, Ph.D.
Case Western Reserve University
Song Yang, Ph.D.
Fred Hutchinson Cancer Research Center
Lue Ping Zhao, M.D.
Fred Hutchinson Cancer Research Center
Argyrios Ziogas, Ph.D.
University of California, Irvine

Contents

Introduction
Daniela Seminara

I. GENE DISCOVERY PANEL

Discovery of Cancer Susceptibility Genes: Study Designs, Analytic Approaches, and Trends in Technology 1
Daniel J. Schaid, Ken Buetow, Daniel E. Weeks, Ellen Wijsman, Sun-Wei Guo, Jurg Ott, Carol Dahl

II. GENE CHARACTERIZATION PANEL

Design of Gene Characterization Studies: an Overview 17
Duncan C. Thomas

Case–Control Studies of Common Alleles and Environmental Factors 25
Neil Caporaso, Nathaniel Rothman, Sholom Wacholder

Family-Based Association Studies 31
W. James Gauderman, John S. Witte, Duncan C. Thomas

Cohort Studies for Characterizing Measured Genes 39
Bryan Langholz, Nathaniel Rothman, Sholom Wacholder, Duncan C. Thomas

Multistage Sampling for Disease Family Registries 43
Kimberly D. Siegmund, Alice S. Whittemore, Duncan C. Thomas

Detection of Interaction Involving Identified Genes: Available Study Designs 49
Alisa M. Goldstein, Nadine Andrieu

Kin-Cohort Designs for Gene Characterization 55
Mitchell H. Gail, David Pee, Raymond Carroll

III. INTEGRATION PANEL

Study Design in Genetic Epidemiology: Theoretical and Practical Considerations 61
Alice S. Whittemore, Lorene M. Nelson

Integrated Designs for Gene Discovery and Characterization 71
Lue Ping Zhao, Corinne Aragaki, Li Hsu, John Potter, Robert Elston, Kathleen E. Malone, Janet R. Daling, Ross Prentice

IV. APPLICATIONS PANEL

Confirmation of Prostate Cancer Susceptibility Genes Using High-Risk Families 81
Gail P. Jarvik, Janet L. Stanford, Ellen L. Goode, Richard McIndoe, Suzanne Kolb, Mark Gibbs, Leroy Hood, Elaine A. Ostrander

Study-Design Issues in the Development of the University of Southern California Consortium's Colorectal Cancer Family Registry 89
Robert W. Haile, Kimberly D. Siegmund, W. James Gauderman, Duncan C. Thomas

Design and Analysis Issues in a Population-Based, Case–Control–Family Study of the Genetic Epidemiology of Breast Cancer and the Co-operative Family Registry for Breast Cancer Studies (CFRBCS) 95
John L. Hopper, Georgia Chenevix-Trench, Damien J. Jolley, Gillian S. Dite, Mark A. Jenkins, Deon J. Venter, Margaret R. E. McCredie, Graham G. Giles

Combined Influence of Genetic and Dietary Factors on Colorectal Cancer Incidence in Japanese Americans 101
Loïc Le Marchand

Introduction

Daniela Seminara

Past genetics and epidemiology research has suggested that cancer has a complex etiology, and that both susceptibility genes and environmental risk factors play a defined role in its development. More recently, this dichotomous theory has yielded to complex, interactive models, involving the temporal and metabolic interplay of both sets of causative agents. This novel conceptual framework entails the disciplinary integration necessary for the identification and characterization of cancer susceptibility genes, the definition of their interactions with nongenetic risk factors during the process of cancer development, and the application of this knowledge to the assessment of cancer risk in diverse populations.

In addition to the conceptual integration from related but traditionally separated disciplines, studying the genetic and environmental causative interaction that determines cancer susceptibility in individuals, families, and populations involves considerable operational challenges, such as the selection (or *de novo* development) of study designs appropriate to the question(s) proposed, the integration of complex models into pertinent analytic strategies, and the interdisciplinary synergy to support and conduct such studies in connection with a rapidly evolving methodologic and scientific milieu. For example, genetic studies on cancer susceptibility usually focus on families at extreme high risk who are selected based on "unusual" family history, while epidemiologic studies of candidate genes or of gene–environment interactions mostly adopt population-based, case–control or cohort designs. The multifaceted nature of a relative young discipline, genetic epidemiology, provides the conceptual framework for the integration of these divergent approaches and for the development of innovative study designs and analytic strategies targeted to further our understanding of the complex mechanisms underlying cancer susceptibility.

This monograph issue presents the proceedings of the conference: "Innovative Study Designs and Analytic Approaches to the Genetic Epidemiology of Cancer," held in Seattle, WA, on May 27–29, 1998, and sponsored by the National Cancer Institute (NCI) and the Fred Hutchinson Cancer Research Center, Seattle. A major goal of this conference was to bring together investigators interested in the development of study designs and new analytic methods pertinent to the genetic and epidemiologic study of complex diseases such as cancer within the framework of modern genetic epidemiology.

The focus of the papers in this monograph is to discuss the usefulness and practical feasibility, potential advantages, and pitfalls of the most recently developed models for study designs and analytic methods in the genetic epidemiology of cancer and to identify analytical and design issues brought to attention by the recent conceptual and technologic advances in the disciplines merging into genetic epidemiology, as related to their application to the following: 1) the pursuit of new cancer susceptibility genes across the expected spectrum of population distribution and penetrance (Gene Discovery Panel); 2) the ascertainment of the population frequency, the phenotype-genotype correlation, and the modifiers of penetrance, such as other genes or nongenetic risk factors, of each identified cancer susceptibility gene (Gene Characterization Panel); 3) the integration of such methodologies and approaches (Integration Panel); and 4) the evaluation of applications of the discussed approaches in currently ongoing studies (Application Panel). The workshop proceedings encompass the results of the discussion of the four interrelated panels.

Although an increasing number of cancer susceptibility genes have been successfully identified during the past few years, design and analytic issues, such as the accurate definition and measurement of cancer phenotype, the use of intermediate end points, the choice of appropriate samples and analytic methods, and the determination of the effect of gene–environment interaction on linkage analysis, still stand in the way of an accelerated rate of discovery. The Gene Discovery Panel, chaired by Dr. Schaid and Dr. Buetow, focuses its discussion on a review of the current designs and analytic challenges in the discovery (mapping) of susceptibility genes for complex diseases such as cancer. The panelists identify the conceptual and methodologic issues representing a barrier to this endeavor; outline a vision for some possible methodologic solutions and for future research in this area; and, finally, examine the conceivable integration of recent and imminent advances in molecular technology into genetic epidemiology studies and its potential for revolutionizing the pace and direction of gene discovery.

The collection of manuscripts from the Gene Characterization Panel, led by Dr. Thomas, addresses the methodologic issues pertinent to studies that assess the population characteristics of cloned genes, such as their allele frequencies, penetrance, variation in these parameters across population, and gene–gene and gene–environment interactions. Dr. Thomas provides an overview of the various designs that have been recently proposed and their appropriateness for addressing specific research questions in the area of genetic characterization, and discusses statistical and practical considerations pertinent to the evaluation of cohort and case–control designs using independent and related individuals, optimal multistage sampling, and hybrid designs. The answer to this question clearly depends on such context issues as nature of the outcome variable, the gene frequency and genetic relative risk, and the importance of gene–environment and gene–gene interactions.

Dr. Caporaso, Dr. Rothman, and Dr. Wacholder propose an novel approach, based on adjustment for ancestry, attention to study design and appropriate selection of controls, and reduction of the confounding caused by population stratification in the analysis of the associations of genetic polymorphisms to cancer outcome in population studies. Dr. Gauderman, Dr. Witte, and Dr. Thomas review case–control designs for studying gene associations in which relatives of cases are used as controls as an

Affiliation of author: Epidemiology and Genetics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.

Correspondence to: Daniela Seminara, Ph.D., M.P.H., National Institutes of Health, Executive Plaza South, Rm. 214, MSC-7147, Bethesda, MD 20892–7147.

alternative strategy to avoid the problem of population stratification that can lead to false associations with noncausal genes. Various case–relative combinations are discussed in light of their validity to estimate genetic relative risk and their efficiency in respect to the more generally used population-based, case–control design using unrelated controls.

During the last few decades, several large cohorts have been established to examine epidemiologic risk factors believed to be associated with cancer. Such cohorts currently include a total of millions of individuals, with extensive baseline and follow-up data, some collection of biologic specimens, and an established infrastructure that could be used to conduct cohort, nested case–control, or case–cohort studies to characterize measured genes in an extremely cost-effective fashion. Dr. Langholz discusses methods of sampling in an efficient manner within these cohorts, the merits of including family members of the original cohort, and the use of molecular markers to control for population stratification. Dr. Siegmund describes the basic statistical theory underlying multistage sampling, a design strategy that can be used to establish family-based disease registries in a cost-effective manner, while retaining the capability of validly addressing a range of scientific questions, from characterizing measured genetic factors and gene–environment interactions to detecting novel susceptibility genes. The application of this theory to a three-stage sampling scheme utilized by the Cooperative Family Registry for epidemiologic studies of colorectal cancer is also described.

Advances in molecular genetic techniques and their integration into epidemiologic designs have led to an increased ability to examine gene–environment interactions. Less studied so far, but with increased potential for the future, is interaction between identified genes. Dr. Goldstein and Dr. Andrieu review the known methods to detect gene–environment and gene–gene interactions in view of the available risk estimates, required types of subjects, feasibility of the proposed study design, and efficiency and power considerations. Further investigations are recommended to define the efficiency spectra of each method. Finally, the advantages and disadvantages of the kin-cohort design, recently developed to estimate the penetrance of BRCA1 in a selected population of Ashkenazi Jewish women, are reviewed by Dr. Gail and colleagues. The potential limitations of this design are examined, as well as the considerable advantages. In particular, the effects of violations of assumptions on estimates of penetrance, the selection bias from preferential sampling of probands with heavily affected families, the misclassification of the disease status of relatives, and the possible violation of Hardy-Weinberg equilibrium are discussed.

The primary objectives of the Integration Panel are as follows: to address the feasibility of integrating genetic and epidemiologic principles in both gene discovery and gene characterization studies; to evaluate strengths and weaknesses of designs used in gene discovery and characterization from the perspective of such integration; and to evaluate strengths and weaknesses of several proposed study designs for interdisciplinary studies in the genetic epidemiology of cancer.

With this goal in mind, Dr. Whittemore and Dr. Nelson review some of the issues that must be considered by genetic epidemiologists when designing a study involving the genetic epidemiology of late-onset chronic diseases such as cancer. The authors discuss how the choice of design is influenced by the study's goals, such as identifying new genes, estimating characteristics of known genes, and learning how to prevent the disease in the genetically susceptible, and propose guidelines for choosing designs that effectively address such goals. Furthermore, the practical constraints in conducting such research are concisely outlined, including problems of potential selection bias, reduced response rates, departures from Hardy-Weinberg frequencies for genotypes, problems particular to family registries, problems particular to the cultures of various ethnic groups, and ethical issues. Last, Dr. Zhao and colleagues expand the challenge to genetic epidemiology methodology by proposing the development of integrated study designs, such as case–family and case–control–family, that would allow for both discovery and characterization of cancer susceptibility genes.

Finally, a series of studies illustrating the practical issues encountered in the implementation of the design and methodologic approaches discussed by the previous panels, and their resolution are presented by several investigators involved in gene discovery and characterization projects. One of the strongest risk factors for prostate cancer identified to date is a family history of disease, suggesting that genetic factors play a role in prostate cancer susceptibility and/or etiology. Dr. Jarvik, Dr. Ostrander, and colleagues review the recent findings suggesting there is a prostate cancer susceptibility gene at 1q24–25 and discuss the results of their gene discovery study in high-risk families and the ones from other data sets with regard to observed difficulties in confirming these initial findings of linkage. Dr. Haile and colleagues describe the implementation of a multistage family-based design performed through the University of Southern California Consortium, Los Angeles (USCC), a participating center in the National Cancer Institute's Cooperative Family Registry for Colorectal Cancer Studies. The primary goals of the USCC are to estimate gene frequency and penetrance and to investigate factors that may affect penetrance (gene–gene and gene–environment interactions) for accepted and putative candidate colorectal cancer genes. Within this general framework, the authors address questions related to the proportions of single versus multiple-case families to include in the registry, to the selection of subjects within families on whom to collect risk factor questionnaires, blood samples, and tumor blocks; and the meaningful selection of samples to test for microsatellite instability and from which to establish cell lines. A protocol to tailor decisions about sampling and data collection to the family structure of each family in an attempt to maximize power for planned analyses per unit cost/effort while retaining a valid design is also described. Last, the issue concerning the appropriate use of data derived from high-risk clinics is addressed.

Dr. Hopper and colleagues report on a population-based study of breast cancer with a case–control family design in which families are identified either through recently diagnosed cases reported to the Victorian and New South Wales Cancer Registries or through controls selected at random from the Electoral Rolls. This design makes it possible to measure the population burden associated with cancer genes and to study concurrently the roles of both genetic and environmental risk factors on cancer susceptibility. Some of the design issues faced in this study as well as the implications of the population-based finding for genetic testing and public health are discussed. The extension of this model to a collaboration with the NCI-supported Cooperative Family Registry for Breast Cancer Studies, an international infrastructure for studies of the genetic epidemiology of

breast cancer, and the practical issues arising out of the first 2 years of experience in establishing and conducting such a novel and innovative project in genetic epidemiology are also reviewed, in view of the benefit of such comprehensive multicenter interdisciplinary infrastructure to research in genetics (with an immediate translation to the population and public health perspective) and genetic epidemiology (allowing for the possible role of genetic factors in explaining the epidemiologic data). Finally, Dr. Le Marchand and colleagues describe recent case–control–family, case–control, and prospective studies of colorectal cancer in Japanese migrants to Hawaii, a population that offered an unique opportunity to identify gene–environment interactions of considerable public health impact, given the common at-risk genotypes and exposures.

Understanding the genetic and environmental determinants of cancer and their interplay at the different causative steps of the etiologic process has tremendous public health relevance, since it constitutes the underpinning of targeted preventive measures, early cancer detection, and effective cancer treatment. It is hoped that the contribution of the scientists participating in this workshop and monograph will further the scientific progress toward this goal. The organizers would like to thank the speakers and the panel chairs for their participation in this effort. Their thoughtful outlooks have helped to identify the emerging and, at times, problematic methodologic and design issues in this important area of research, to further a discussion on resolving such issues, and to transmit this knowledge to the scientific community at large.

We also thank the reviewers of the published manuscripts for making this monograph possible.

I. GENE DISCOVERY PANEL

Discovery of Cancer Susceptibility Genes: Study Designs, Analytic Approaches, and Trends in Technology

Daniel J. Schaid, Ken Buetow, Daniel E. Weeks, Ellen Wijsman, Sun-Wei Guo, Jurg Ott, Carol Dahl

Determining the genetic causes of cancers has immense public health benefits, ranging from prevention to earlier detection and treatment of disease. Although a number of cancer susceptibility genes have been successfully identified, design and analytic issues remain that challenge the current paradigm of gene discovery. Some examples are the definition and measurement of cancer phenotype, the use of intermediate end points, the choice of sample (e.g., affected relative pairs versus large extended pedigrees), the choice of analytic method [e.g., parametric logarithm of the odds (LOD) score method versus model-free methods], and the influence of gene–environment interaction on linkage analysis. Furthermore, association methods, based on either the traditional case–control study design or family-based controls, are popular choices to evaluate candidate genes or screen for linkage disequilibrium. Finally, the study design and analytic methods for gene discovery are determined to some extent by what genomic technology is feasible within the laboratory. Many of the main issues related to gene discovery, as well as trends in genomic technology that will impact on gene discovery, are discussed from the perspective of their strengths and weaknesses, pointing to areas in need of further work. [Monogr Natl Cancer Inst 1999;26:1–16]

Determining the genetic causes of cancers has immense public health benefits, ranging from prevention to earlier detection and treatment of disease. Numerous studies have demonstrated that both somatic genetic changes and hereditary factors are involved in the etiology of many cancers. Although somatic genetic changes may appear to be more frequently related to cancers than hereditary factors, the primary focus of this paper is on the discovery of hereditary factors—those genes that increase the susceptibility to cancer. Genetic mapping consists of two different yet interrelated problems: to estimate the genetic effects and to localize the gene(s). The biggest challenge today for gene discovery is to find the locations of the genes with little or no knowledge at all of how many genes are involved, how they interact with each other or with environmental factors, and what the genotype–phenotype relationship is. Although the contribution of a single gene may account for only a portion of susceptibility to the disorder, the common occurrence of the disorder warrants a better understanding of its causes. Never before have the available genetic tools, and those on the near horizon, offered as much promise to help in this determination as they offer today. The promises are many, but so too are the complexities of common cancers. For example, some complexities of heritable disorders are incomplete penetrance (<100% risk of disease among carriers), phenocopies (disease among noncarriers), locus and allelic heterogeneity, oligogenic and polygenic inheritance, gene–environment interactions, gene–gene interactions, mitochondrial inheritance, parental imprinting (differential gene expression depending on the sex of the transmitting parent), and anticipation (a trait that progressively increases in severity in successive generations, often via expansion of trinucleotide repeats). The hallmarks of single gene disorders are a very high recurrence risk compared with the general population, with the risk to relatives decreasing by a factor of one half with each degree of relationship, clustering of disease within families that tends to follow predictable mendelian transmission patterns and rare mutant disease-causing allele(s). In contrast, many common diseases with a complex genetic basis do not demonstrate the simple mendelian expectations when multiple genes interact to increase the disease risk, and the recurrence risk to relatives decreases with each degree of relationship more rapidly than a factor of one half *(1)*. Of course, there are also many complex traits that do not demonstrate mendelian expectations that still maintain a decline in recurrence risk of about one half with each degree of relationship. These complexities emphasize the need for efficient study designs and analytic methods for gene discovery.

The goals of this paper are the following: 1) to present the main design and analytic approaches that are currently available for the discovery of disease susceptibility genes; 2) to highlight some of the complex, and perhaps debatable, issues regarding gene discovery; 3) to offer directions for future methodologic research; and 4) to discuss recent genetic technological advances that will influence the future directions of gene discovery.

Definition and Measurement of Cancer Phenotype

Some of the clinical clues that there may be an inherited predisposition to cancer are 1) the occurrence of cancer at an unusually young age, relative to the typical age for the type of cancer; 2) multifocal development of cancer in a single organ or bilateral development of cancer in paired organs; 3) development of more than one primary tumor of any type in a single person; 4) family history of cancer of the same type in close relative(s); 5) high rate of cancer in a family; and 6) the co-occurrence of congenital anomalies or birth defects in a person with cancer or within a cancer-prone family *(2)*.

Affiliations of authors: D. J. Schaid, Departments of Health Sciences Research and Medical Genetics, Mayo Clinic/Mayo Foundation, Rochester, MN; K. Buetow, C. Dahl, National Cancer Institute, National Institutes of Health, Bethesda, MD; D. E. Weeks, Department of Human Genetics, University of Pittsburgh, PA; E. Wijsman, Medical Genetics and Biostatistics, University of Washington, Seattle; S.-W. Guo, Division of Epidemiology, University of Minnesota, Minneapolis; J. Ott, Rockefeller University, New York, NY.

Correspondence to: Daniel J. Schaid, Ph.D., Harwick 7, Mayo Clinic, 200 First St. SW, Rochester, MN 55905 (e-mail: schaid@mayo.edu).

© Oxford University Press

The definition and measurement of the cancer phenotype are critical for gene discovery because misclassification of phenotype status can lead to loss in power for familial aggregation studies as well as for linkage and association studies. The level of certainty of cancer diagnosis will likely differ for different types of studies. For example, it may be most feasible to collect family history of disease by a single family member (informant) or perhaps supplement this information with a second family informant, when performing large studies of the aggregation of disease within families or performing complex segregation analyses. The accuracy of the primary site of cancer based on family history has been reported to be high for first-degree relatives (85% accurate) and lower for second-degree relatives (67% accurate) *(3,4)*, suggesting that more effort is needed to validate cancers among more distant relatives. For example, a single informant within each nuclear family of a pedigree may be a reasonable strategy. Furthermore, the accuracy tends to vary according to the site of the primary cancer, with lower accuracy for cancers in female pelvic organs (e.g., ovary, cervix, and uterus), and in sites where metastasis is common (e.g., liver and bone). These inaccuracies will bias the estimates of genetic parameters if not accounted for. A method to account for measurement errors is to randomly sample a validation sample from the total of all subjects, validate the cancer phenotype status of all subjects in the validation sample, and use the validation sample to estimate the probability of the true disease status, given the disease status reported by family history. This probability distribution can then be used to correct those phenotypes that were not validated, by including this correction probability distribution in the statistical analysis. Similar methods have been proposed to account for measurement error of covariate data *(5)*. Although this approach is straightforward to implement in the design of a study, current genetic analysis software does not implement this type of validation information. For linkage and association studies of either candidate genes or genome-wide scans, it is generally accepted that it is necessary to confirm, by medical records (preferably pathology reports), the cancer diagnoses for all subjects included in the analyses. Given the efforts required to collect blood specimens and perform the genotyping, and the consequences of misclassification on gene discovery, this viewpoint is warranted.

Because genes may be related to multiple outcomes, such as multiple cancer types (i.e., pleiotropy), it is not always clear what the best definition of cancer "phenotype" should be. Most current linkage studies focus on a single cancer type, such as breast cancer or prostate cancer. Although this may be adequate for rare genes of high penetrance that mainly affect a single cancer site, this approach can sacrifice power for common genes with low penetrance that affect multiple cancer sites. Further work, along the lines of characterizing the relationship between genotype and phenotype, is needed to better understand the best methods to discover these types of susceptibility genes. Further clues in the discovery of cancer genes can be gathered by looking for traits that could suggest chromosomal rearrangements, such as mental retardation or birth defects.

Intermediate End Points

Other traits that may be intermediate end points for cancer may be worth considering for use in gene discovery studies. The rationale for considering intermediate end points is that, being closest to the responsible genetic components that eventually lead to the disease of interest, the intermediate end points can thus be easier to map in a more efficient way. Of course, a potential flaw with this approach would occur when the intermediate end points are actually more remotely determined by the genetic components responsible for the trait of interest, leading to a loss in efficiency. Because quantitative traits may be useful for linkage analyses and are oftentimes more powerful than binary traits derived from quantitative traits *(6)*, it is tempting to use intermediate traits as surrogates for cancer diagnoses. Some examples of such traits are breast density (a risk factor for breast cancer); counts of nevi (related to malignant melanoma); polyps and precancerous pathologic features, such as colonic aberrant crypt foci (related to colon cancer); and prostate-specific antigen (related to prostate cancer). Statistical methods to use surrogate and auxiliary end points have been developed for clinical trials in which the true end point of interest is the time to failure. In the context of gene discovery, age at cancer onset can be substituted for "time to failure," so that many of the same principles for surrogate end points can be applied to designs and analyses for the discovery of genes that increase the risk for early age at onset of cancer. According to the principle laid out by Prentice *(7)*, a surrogate end point is used to define a trait for which a test of the null hypothesis of no relationship with a genetic marker or a candidate gene is also a valid test of the corresponding null hypothesis based on the true end point. The principal criterion for a surrogate trait is that it fully explains the relationship between the genetic marker and the cancer phenotype. Statistically, this requires that the distribution of the true end point does not depend on the genetic marker, given the value of the surrogate trait. In other words, the relationship between the true end point and the surrogate trait is the same for all marker genotypes. To emphasize this by example, consider a gene (G) related to colon cancer (C), and the use of colonic polyps (P) as a surrogate trait. If the gene causes polyps, and then polyps lead to an increased risk of colon cancer, then P would be a valid surrogate if the risk of colon cancer does not depend on G, once P is measured. This can be illustrated from a mechanistic viewpoint by arrows leading from the cause to effect as illustrated in Fig. 1.

In contrast, P would not be a completely valid surrogate if the risk for colon cancer depends on knowing both P and G, with an example illustrated by the causal pathway in Fig. 2.

Fig. 1. Example of a causal path from a genetic cause (G), to the intermediate trait polyps (P), to the end point colon cancer (C).

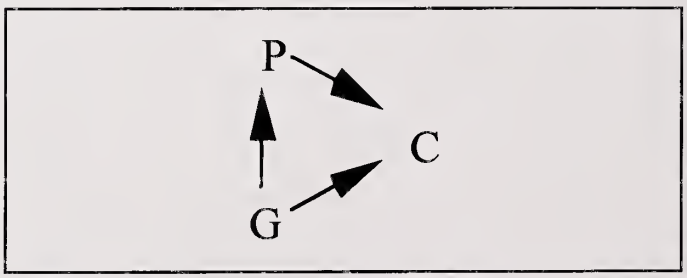


Fig. 2. Example of a mechanism by which the gene (G) influences the risk of polyps (P), yet, even after adjustment for the influence of polyps on the risk of colon cancer (C), the gene still has an influence on the risk of colon cancer.

Although Prentice's operational criterion for a valid surrogate end point is strict, and the second example above may still offer some information on the association of the genetic marker with cancer phenotype when using polyps as a surrogate, the criterion does offer guidance when evaluating a potential surrogate trait. So, before using a trait as a surrogate for cancer phenotype, there should be evidence that the surrogate is heritable and that it is appropriately related to the outcome of interest. To appropriately use a surrogate trait, it is best to obtain a validation sample on a subset of subjects, so that the validation subset can be used to statistically relate the surrogate outcome with the cancer phenotype *(8)*.

Sampling Considerations

The power of a linkage study to detect a susceptibility gene depends on the frequency of informative families and the ability to use informative markers. A basic principle to consider when planning a study is that recombinant meioses can be identified only among subjects who are doubly heterozygous for both the disease locus and the marker locus. Because the underlying genotype is only indirectly observed through the phenotype, ambiguity of whether the underlying genotype is doubly heterozygous and ambiguity with regard to the linkage phase of the disease and marker loci both diminish the amount of linkage information. When penetrance is low, when phenocopies are frequent, or when multiple interacting genes are involved, the relationship between phenotype and genotype becomes more ambiguous, leading to increased sample size requirements. Sample size requirements can be reduced by increasing the informativity of the markers, by using markers that are in linkage disequilibrium with the disease locus, and by extending pedigrees to include informative relatives. An informative relative is one who helps to determine whether a person in the pedigree is a double heterozygote as well as the phase of the double heterozygote. This determination suggests that including grandparents who have discordant phenotypes should increase the amount of linkage information *(9)*. Another way to increase linkage information is to restrict analyses to clinical subsets that demonstrate mendelian inheritance. For some cancers, such as breast and colon cancers, sampling pedigrees that demonstrate mendelian inheritance patterns has led to successful gene discovery (e.g., BRCA1 and BRCA2 for breast cancer).

Although it is difficult to give general sampling guidance without considering the specifics of the potential genetic mechanisms underlying a trait, some guiding principles have been provided by Risch *(10)* and Weeks *(11)*. For a mendelian autosomal recessive disease, nuclear families with unaffected parents and many affected offspring are informative (because both parents are heterozygous for the susceptibility gene), as are inbred families *(12)*. For a rare mendelian autosomal dominant disease, extended multigenerational pedigrees with many affected subjects are especially informative.

Some of the advantages of large pedigrees over nuclear families are that there can be increased power to detect linkage (*as outlined above*), and genetic homogeneity is likely to be increased. However, there are limitations of large pedigrees that nuclear families or sib-pairs can overcome: Specific mutations in rare large pedigrees are probably different from those that cause the more common form of the disease; for common diseases, large pedigrees do not necessarily guarantee genetic homogeneity, and phenocopies are more likely to occur, which can dilute the power for linkage; rare large, dense pedigrees for common diseases may represent a chance combination of multiple genetic and/or environmental factors, which can diminish the chance of detecting a single gene by linkage methods; multigeneration pedigrees are likely to bias the detectable genetic susceptibility gene to a dominant mode of inheritance; errors in the phenotype of the disease or marker in a few key people in a large pedigree can have profound influences on the linkage results *(13,14)*. When the population-attributable risk of a susceptibility gene is assessed, a large number of families is required to reliably estimate the fraction of families that show linkage of disease to the marker, which is feasible when sampling nuclear families.

In contrast, under oligogenic epistatic models, which may be plausible for some complex traits *(1)*, affected sib pairs (ASPs) are more likely to be informative for linkage than are large pedigrees with multiple affected members *(10)*. This point can be understood as follows: If the genetic mechanism involves several interacting genes of small to modest effects, and large extended pedigrees with many affected subjects are sampled, then parents who have many affected children can have an increased probability of being homozygous for the disease locus, especially if the parents are also affected, causing the parents to be uninformative for linkage. Also, if multiple susceptibility loci are involved, and susceptibility alleles at these loci are fairly common, then a pedigree with many affected subjects has an increased chance of having multiple susceptibility alleles segregating, diminishing the power to detect any one of the loci. Without strong evidence that mendelian inheritance occurs for the disease of interest, a popular option is to sample ASPs, or other affected relative pairs, and use model-free methods of analysis that do not require specification of the penetrance and mode of inheritance. A popular sampling consideration for ASPs is to sample families enriched for a large number of affected sibs and to sample according to the parents' affection status, with preference for unaffected parents. However, the power of these sampling criteria depends on the underlying genetic mechanism; for some genetic models, particularly two-locus epistatic models, sampling ASPs from sparsely affected sibships can offer greater power than sampling from densely affected sibships. That is, multiple affected sibs increase the probability that additional susceptibility alleles are segregating in the family, causing a reduction in the average power of each sib pair. An alternative sampling scheme is to simply sample all ASPs, regardless of the affection status of first-degree relatives, and recognize that the potential loss in power of this strategy, which may be only modest for some genetic models, can be compensated by increasing the sample size. These points, and further heuristics for ASP sampling, are discussed by McCarthy et al. *(15)*. It is noteworthy that the ASP design for linkage remains a controversial issue. Results from the Genetics Analysis Workshop 10 indicate that, for traits with low epistatic and dominance variance, ASP methods can be very inefficient relative to sampling large pedigrees, even in the absence of a well-defined genetic model *(16)*. A subtle point for multilocus diseases is that ASPs can be particularly poor for searching for additional loci, conditional on the first locus being known *(17)*. Further discussion regarding the relative merits of ASPs versus pedigree data is given by Schork and Xu *(18)*.

Both simulation *(19)* and analytic *(9,20,21)* methods can be used to determine the power of parametric logarithm of the odds (LOD) scores to detect linkage for a specified genetic mechanism, and methods to determine power for ASP methods have been published for both binary affection status *(15,22,23)* and quantitative traits *(24,25)*.

Analytic Methods

Powerful methods of genetic resolution depend on linkage or association of traits, or a combination of both, with genetic markers or candidate genes. A gene may be measured directly at the DNA level or at the level of the gene product as well as indirectly by DNA markers (such as simple sequence repeats). It is convenient to refer to any of these measured phenotypes as a genetic marker, which is defined as a genetically determined trait for which the relationship between genotype and phenotype is known. Linkage analysis of complex diseases has proven to be an effective method to identify genes related to disease subtypes that are primarily monogenic, such as BRCA1 and BRCA2 mutations that cause breast cancer. Linkage has also been useful for diseases that are likely caused by multiple genes, such as insulin-dependent diabetes. Although linkage of insulin-dependent diabetes has been obtained with up to 18 different chromosome regions *(26–28)*, recent reports *(29,30)* have suggested that there may be much fewer regions involved.

Other disease genes will likely be identified with the use of linkage analyses. However, recent advances in molecular genetics and statistical techniques have led to even more powerful tools, using genome-wide association studies *(31,32)*. The potential advantage of association studies over linkage is that any association of a marker with a trait, because of either a pleiotropic effect of the marker (e.g., a candidate gene) or tight linkage with disequilibrium between the marker and trait loci, is much easier to detect than linkage without disequilibrium *(32–37)*. This advantage has historically motivated many association studies of complex diseases with genetic markers and candidate genes, despite the biggest potential pitfall—the choice of control group. Using diseased cases and unrelated control subjects is prone to spurious (i.e., nongenetic) associations because of population stratification, admixture, and migration. For more than 40 years, scientists have recognized that relatives of diseased cases can serve as unbiased control subjects that avoid the potential spurious associations given by unrelated control subjects *(38)*, but only recently have scientists begun to explore the statistical methods needed to design and analyze such family-based association studies. Below we briefly review the main features of linkage methods and association methods, pointing out their strengths and weaknesses to consider where further work is needed.

Linkage Methods

Linkage analysis is one of the basic tools used to evaluate the coinheritance within families of genetic markers with a trait and has proven to be highly successful at mapping monogenic traits as well as some diseases that are likely caused by multiple genes. The classic LOD score method *(39)*, based on likelihood principles for an assumed genetic model, has been the most widely used method for linkage analysis and the most successful in terms of the number of genes subsequently cloned. These successes have been primarily for simple mendelian disorders that are inherently much easier to genetically map than are complex disorders. The theoretical foundation of the LOD score meth is well understood *(40)* and implemented with sophisticated gorithms *(41–46)* in widely available computer software. basic feature of the LOD score method is an assumed ger model for both the trait phenotype and the marker phenoty The genetic model for the trait phenotype requires specifica of autosomal or X-linked trait loci, the number of loci, the nu ber of alleles at each locus and their frequencies, the joint g notype probabilities, and the genotype-specific penetrances, which can vary with a known risk factor (e.g., liability classes). Similar genetic parameters are required for the marker phenotype, but these parameters are most often easier to specify. The difficulty in using the LOD score method for complex traits is specification of the genetic model for the trait phenotype. Although errors in the model can lead to inconsistent parameter estimates and lack of power, incorrect models do not lead to an increased false-positive rate unless the parameters for both the trait and the marker loci are wrong *(47,48)*.

Because of the difficulty in specifying the correct genetic model, allele-sharing methods that do not require this specification ("model-free" methods) have become popular for studying complex traits *(1,22,49–52)*. Sharing of alleles between relatives can be considered for either the number of alleles shared identical by descent (IBD) or identical by state (IBS). Alleles shared IBS may not be of the same ancestral origin, and so this information is confounded between linkage and the population frequency of the marker alleles. Because linkage requires information on inheritance, IBD information offers greater power than does IBS information. Allele-sharing methods are often based on affected relatives, usually ASPs, which simplify data collection. This approach offers a big advantage for complex traits, because large sample sizes are often needed to detect the modest effects of susceptibility genes. Note, however, that collection of unaffected family members can substantially increase the power of ASP methods by increasing the IBD inheritance information *(50)*. There has been immense growth in the number of statistical methods for ASP designs, ranging from use of complete family data to infer genotypes to extract IBD information *(53)*; to consideration of the optimal weights used to combine sibships with varying numbers of affected members *(54)*; to inclusion of different types of affected-relative pairs *(55)*; to improved maximum likelihood IBD sharing estimation and model-free maximum LOD scores (MLS) *(22)*; to interval mapping and exclusion of chromosomal regions based on recurrence risk ratios for complex diseases *(56)*; to improved likelihood methods by restriction to genetically plausible IBD sharing ("possible-triangle" constraints) *(57)*; to two-locus disease models that simultaneously evaluate two marker loci, leading to improved power to detect oligogenic disease mechanisms *(58–60)*; to complete multipoint extraction of IBD information for small pedigrees *(46)* and improved nonparametric scoring functions *(61)*. The GENEHUNTER software (Whitehead Institute for Biomedical Research, Boston, MA) *(46)* incorporates these latter two features, but the proposed linkage test tends to be conservative when the IBD information is incomplete, which often occurs. However, a recent extension to GENEHUNTER, based on inclusion of only one additional parameter for the scoring function *(40)*, can dramatically improve the accuracy of the linkage test *(62)*. A recent comparison of 23 different statistics for model-free linkage analysis of nuclear family data indicated that the ASP "mean" test *(50)*, which compares the observed number

alleles shared IBD with its null expected value, performs well terms of type-I error and power for a variety of genetic effects; ension of this test to include unaffected sibs increases power . Of interest, this mean test has been shown to be the uni- nly most powerful test under a recessive mode of inheritance is locally optimal otherwise *(63)*. Extension of the idea that odel-free score tests are optimal for a genetic model *(40)* al- ws a general framework in which powerful statistical tests can be developed.

The above methods are generally based on IBD information. Alternatively, the affected-pedigree-member (APM) method of linkage analysis is a nonparametric statistic that measures the similarity of marker alleles shared IBS among affected relatives *(49)*. This approach is computationally efficient *(64)* and has been extended to include unaffected relatives *(65)*, X-linkage *(66)*, multipoint analyses *(67)*, and linkage heterogeneity *(68)*. The APM method has been advocated as an alternative to IBD methods because only affected members are required, which is convenient for situations in which it is difficult to determine IBD information, such as for late-onset disease *(69)*. But there is a sacrifice in power compared with methods based on IBD status *(70–72)*. One of the greatest drawbacks of the APM method is that it is not robust (i.e., inflated false-positive rates) to misspecified marker allele frequencies *(73)*. However, a permutational extension of the APM method, called the SimIBD statistic *(72)*, uses IBD information when it is available and IBS when it is not. The SimIBD statistic can be markedly less sensitive to misspecification of marker allele frequencies than APM, because the statistic is computed conditionally on the marker genotypes of the nonaffected relatives.

Another nonparametric linkage method, the weighted rank pairwise correlation (WRPC) statistic *(74–76)*, is also based on IBS information. But, unlike the APM method, the WRPC statistic allows for not only binary traits but also quantitative and age-dependent traits as well as inclusion of nongenetic covariates.

Finding genes in the presence of strong environmental risk factors may be difficult, so control (at least in the statistical sense) of environmental risk factors may be crucial. Although parametric linkage analyses can allow for environmental strata, by allowing the penetrance of the susceptibility gene to vary across environmental strata (i.e., liability classes), it is difficult to implement liability classes in practice because of unknown penetrance values. Further statistical work is needed to appropriately account for measured environmental factors. Because some environmental risk factors may have large measurement errors (such as diet surveys), it may be worthwhile to obtain a validation sample on a subset of subjects so that the validation subset can be used to account for the measurement errors.

When considering covariates, researchers should also consider the best methods of design or analysis to account for gene–environment interaction when searching for genes. Finding genes that interact with environmental factors will be a great challenge. Virtually all complex traits are known, or suspected, to be influenced by various environmental risk factors. And many of them are believed to be influenced by both genetic and environmental factors and their interactions. However, very little is known about how genetic and environment factors interact and about how this interaction affects current mapping strategies. In fact, there is a lack of a precise definition of gene–environment interaction and a lack of appreciation of the fact that there is a distinction between statistical interactions and biologic interactions.

Researchers now know that, in the presence of gene–environment interaction, the widely used measure of familial aggregation, the recurrence risk ratio, can be a poor measure for the strength of genetic effects of the underlying genetic component, simply because the genetic and environmental factors are now confounded. Other measures, such as heritability and twin concordance rates, suffer the same problem. In addition, a single-locus model with environment factors can display surprisingly complex behaviors that are otherwise unexplainable under a pure, single-locus model without environment factors. Moreover, at least for sib-pair designs, the presence of gene–environment interaction can markedly affect the statistical power to detect either linkage or association.

Despite the many advances in the development of nonparametric statistics for linkage, debate still exists regarding use of the classic LOD score method versus model-free methods for complex traits *(77–80)*. By model free, we mean that the analytic method does not require specification of the frequency of the susceptibility allele or the mode of inheritance and penetrance. Technically, even the model-free methods require some basic genetic assumptions, such as mendelian transmission (as opposed to meiotic drive) and no interference. However, one should be aware that the power of model-free methods depends on the underlying genetic mechanisms, and so the power of model-free methods can be improved if the genetic basis of the disease being studied matches the assumptions of the model *(40)*. Analytic results indicate that the LOD score method is robust to misspecification of either the genetic parameters related to the trait, or those related to the markers, but not misspecification of both sets of parameters *(47)*. If ascertainment is through either the marker or trait phenotypes but not both, then the LOD score does not depend on the mode of ascertainment. For complex diseases, the genetic parameters related to the trait are rarely known, so it is critical to have accurate marker allele frequencies to avoid an inflated chance of a false-positive finding. But correct allele frequencies are not sufficient if the analyzed sample represents an admixed or stratified population, because information on the population structure must be included to fully describe the distribution of the marker genotypes. Unfortunately, because this population information is rarely known, it is often ignored in linkage analyses. Ignoring population stratification in linkage analyses can lead to an inflated false-positive rate, as demonstrated for ASP analyses *(56)*.

When the genetic parameters are misspecified, it is not surprising that there is some loss in power of the LOD score method *(35)*. The amount of power loss depends critically on the mode of inheritance and, to a lesser extent, on the allele frequencies, on the penetrance values, and on the size of the pedigree *(40)*. Simulations suggest that, if two loci epistatically influence disease, accounting for the second locus can increase power *(81)*. If, however, there is locus heterogeneity, and a linked locus accounts for less than 25% of the diseased families, then the power of both the LOD score admixture method, which allows for a fraction of unlinked families *(82,83)*, and the ASP method *(84)* is so weak that the sample size requirements are prohibitive.

Artificial Neural Networks and Linkage

For complex traits, it is generally believed that multiple, possibly interacting loci confer susceptibility rather than a single-

disease locus. Therefore, pattern recognition techniques have been developed with the aim to identify sets of marker loci that jointly show deviations from random allele sharing. Some of the markers may act only as modifiers for others. For example, allele sharing may be increased at one marker only when allele sharing is elevated at another marker. If there are interactions among disease loci, pattern recognition techniques may be able to recognize these interactions and, therefore, be more powerful than conventional ASP methods.

Pattern recognition techniques are often used to predict an outcome (e.g., disease diagnosis) on the basis of observed predictor variables. This prediction is achieved by monitoring observations with known predictor variables and outcomes. Linear discriminant analysis may then be applied to determine the set of predictor variables that best predict outcome. For allele-sharing data in ASPs, a particular type of pattern recognition technique called artificial neural network (ANN) can be employed, which is able to carry out nonlinear discriminant analysis *(85)*. The network architecture is a feed-forward ANN with three neuronal layers (input, hidden, and output).

ANNs were originally developed as simple models for the way nerve cells transmit impulses in the brain. In ANNs, neurons are modeled as nodes that are arranged in layers. A given node potentially receives impulses from all nodes in the layer preceding it and potentially sends an impulse ("fires") to each node in the layer following it. Whether a node fires is determined by the sum of all impulses received and some threshold function. On the basis of multiple observations at the input nodes and predetermined values of the output nodes, ANNs "learn" how best to connect input and output nodes via pathways through hidden nodes. The "strengths" of connections among pairs of nodes in adjacent layers are called weights, and iteratively estimating these weights is called training a network. A common training (learning) algorithm is the back-propagation (a simple downhill) method. Initially, weights have some assumed starting values, which are modified and optimized in the course of learning. These weights are analogous to the coefficients associated with predictor variables in discriminant or multiple regression analysis. ANNs may be seen to carry out simple calculations in a highly parallel manner, which enables them to do tasks for which other methods are unsatisfactory.

Allele-sharing observations have a very simple structure: For each parent, at any marker, an ASP shows sharing ($x = 1$) or no sharing ($x = -1$). Thus, for all markers studied on the genome, observations form an array (a matrix) of x values with m rows (markers) and n columns (parents). The ANN has m input nodes, a hidden layer with a smaller number of nodes, and an output layer with two nodes, O_1 and O_2. The network is "trained" (estimates the weights) with 1) observed allele-sharing data and 2) data randomly generated on the computer as input data. The rationale is that ASP data contain "signal" (disease loci) and "noise" (random allele sharing at markers unlinked with disease loci), while randomly generated data contain only "noise." Correspondingly, training of the ASP data is done with output nodes set to $O_1 = O_2 = 1$, while for randomly generated data they are set to $O_1 = 0$ and $O_2 = 1$. Then one differentiates between weights to the two output nodes, $O_1 - O_2$ (signal + noise – noise = signal) and computes a compound contribution value, C_i, $i = 1 \ldots m$, for each marker locus.

Application of these methods to published allele-sharing data for a genome screen of diabetes genes *(27)* showed that the ANN recognizes all markers found by conventional methods. In addition, it points to markers not implicated in the original study. In particular, one of the genes (IDDM4) implicated in diabetes was seen in the original genome screen only after sib pairs were subdivided into two groups—those that showed allele sharing at human leukocyte antigen (HLA) for both parents and those that did not, which indicates an interaction between IDDM4 and a diabetes gene (IDDM1) at the HLA region. In contrast, the neural network recognizes IDDM4 without subdividing data.

This neural network approach to gene mapping appears promising for various complex traits, such as cancer. Data types other than ASPs, in which recognizing patterns among observations is important, may also be suitable for analysis by neural nets. For example, many cancer tumors are homozygous for a marker, while the individual is heterozygous. This loss of heterozygosity (LOH) is a well-known method for identifying cancer genes. In a genomic survey for LOH, neural nets may be able to pick out patterns of interactions among cancer-causing genes that are difficult to recognize otherwise.

Association Methods

Association studies have often been an essential step after linkage analyses have indicated linked chromosomal regions because the resolution of linkage studies is, at best, approximately 1–2 centimorgans for simple mendelian diseases and feasible sample sizes *(86)* and is likely to be less refined for complex traits. In contrast, the mapping resolution based on linkage disequilibrium (LD) can be as fine as 50–75 kilobases *(87)*. For simple mendelian diseases, there have been numerous statistical methods proposed for LD fine-scale mapping *(88–95)*. However, for complex diseases, family-based association studies may be more desirable than strictly linkage-based methods because they can have greater power, as shown for ASPs *(31)*. With improved genetic marker maps and maps of expressed sequence tags *(96,97)*, it will be feasible to perform genome-wide association studies *(31)*. Besides family-based studies, other study designs for associations are case–control, cohort, cross-sectional, and admixed populations *(98,99)*. But family-based designs are least prone to biased associations caused by the genetic structure of the population.

Although the causes of association between genetic markers and traits can be difficult to determine by association studies, it is important to design studies and analytic methods that distinguish between genetic and nongenetic causes of association. Two genetic causes of association are 1) the pleiotropic effects of the genetic marker (as in a candidate gene study) and 2) LD of the trait and marker loci. Note that LD occurs when haplotype combinations of alleles at different loci do not occur randomly. This LD can occur when most diseased subjects inherit from a common ancestor a segment of a chromosome containing the disease allele and marker allele (founder effect), and LD is most easily detected in a genetically homogeneous population. In this case, the main factor influencing LD is recombination between the two loci. If the recombination fraction is small, as expected for dense marker maps, the amount of LD will remain large even after many generations, making association studies based on LD a potentially powerful strategy for complex traits.

Other causes of association between a trait and marker allele(s) that are not due to linkage, yet which can mislead interpretations, are joint selection for both marker and trait locus alleles, small population variation (i.e., random genetic drift),

and the structure of the population, which may include inbreeding, admixture, or stratification of different ethnic groups. If a population is composed of a recent admixture of different ethnic groups that have different frequencies of marker alleles, then any trait more frequent in an ethnic group will be positively associated with any marker allele that is more frequent in that group, even if these loci are not linked. However, because disequilibrium as a result of admixture, selection, or drift between unlinked loci decays very rapidly over generations, but slowly for linked loci *(100)*, mapping genes via association remains appealing, especially when it is possible to control for nongenetic associations by using family-based association studies *(101)*.

Although association studies have become a popular and effective method of fine-scale gene localization for simple diseases, the complexities of common diseases will challenge association methods. Diallelic markers and multiallelic markers have been compared in terms of sample sizes required to detect LD in a case–control study for diseases with mendelian inheritance, and these results have been extrapolated to discuss the feasibility of LD screening in more complex situations. Results showed that multiallelic markers always have more power to detect LD than diallelic markers (under otherwise equivalent conditions) and that the ratio of the number of diallelic to multiallelic markers needed for equivalent power increases with mutation age and complexity of mode of inheritance. Increasing complexity has particularly negative effects on power to detect LD, especially in the context of a diallelic genome screen. Such complexity can be introduced by dominance, allelic or locus heterogeneity, or presence of sporadic cases. Equivalent power to that achieved by a multiallelic screen can theoretically be achieved by using a more dense diallelic screen, but mapping panels of the necessary resolution are not currently available and may be difficult to achieve. Genome screening with the use of LD testing may therefore only be feasible for young (<20 generations), rare, monophyletic mendelian diseases, such as may be found in rapidly growing genetic isolates *(102)*. Similar conclusions were drawn for transmission disequilibrium testing with the use of diseased cases and their parents *(103)*.

Because of the difficulty in defining an appropriate control group for association studies in heterogeneous populations, Falk and Rubinstein *(104)* proposed to measure the genetic marker on both the diseased cases and their parents (which we refer to as parental control subjects) to compare the frequencies of those alleles that were transmitted from parents to children versus those that were not transmitted. For this type of analysis, only heterozygous parents are used (homozygous parents are not informative), and the genotype of each parent can be considered a matched pair of alleles, one transmitted and the other not; the McNemar statistic for matched pairs, which does not require independence of parental alleles, leads to a valid statistical test. This is also called the transmission/disequilibrium test, or *TDT* *(105,106)*. An alternative approach, which uses all parental alleles, is to ignore the matching in order to compare the frequency of an allele among the $2n$ transmitted alleles versus the $2n$ nontransmitted alleles, where n is the number of cases [also called the haplotype-based haplotype relative risk statistic, HHRR *(107)*]. However, this method requires independence of parental alleles in the population, which is not true for a stratified population *(105)*.

The statistical properties of various statistics and strategies for *TDT*-type analyses for binary phenotypes have been investigated *(107–122)* with considerations of statistics for multiple marker alleles *(106,123–127)*, exact tests *(128)*, designs for admixed populations *(103,129)*, design issues for incomplete LD *(130)*, unbiased tests for association in the presence of linkage when using multiple affected sibs *(131)*, likelihood methods useful to distinguish positively and negatively associated alleles *(132)*, modeling gene–environment interaction *(133)*, and analyzing marker haplotypes with two loci *(134)*. Also, a general framework based on a conditional likelihood method has been presented *(111,135,136)*, which has allowed development of powerful omnibus score statistics *(137)*. This framework has also been used to compute maximum likelihood estimates of allelic effects and to assess interactions between marker genotypes and environmental covariates *(135,138–140)*.

Because data on parents can be difficult to obtain, especially for late-onset diseases, sib-control subjects offer a valid alternative *(141–148)*, although sib-control subjects have less power to assess transmission disequilibrium than do parental marker genotype data *(147,148)*. It is critical to recognize that designs using parents or sibs as controls are sensitive only to associations caused by both linkage and LD and hence avoid the biases due to nongenetic causes of association.

Complications of associations caused by LD are that different mutations causing the same trait can arise on chromosomes that bear different marker alleles, or mutations at the marker locus can occur. These factors can cause the associated marker allele to differ across different populations and can decrease LD in a population mixed with different mutations. To study this effect in detail requires studies of haplotypes composed of multiple marker loci to determine if particular haplotypes are associated with disease. The statistical methods used to evaluate associations with marker alleles can also be used to evaluate associations with haplotypes created from multiple marker loci. Haplotypes can be inferred either by family studies or statistically by using measures of population LD to predict the most likely haplotypes *(149)* or a combination of both *(150)*.

With the availability of many genetic markers, one of the most challenging statistical issues is the choice of the level of statistical significance to maximize power yet minimize the chance of false positives. When testing many markers, the prior probability that any one is associated with disease is often so low that one needs to be very conservative to avoid false-positive associations. To account for multiple comparisons with many alleles (or many haplotypes), the Bonferonni correction is often used, although the power of this method is often weak. Two-stage designs may prove to be useful *(111)* as well as *P*-value plots *(151)* and empirical Bayes shrinkage estimates *(152,153)*.

Association studies using more extensive pedigree data offer yet another useful design because members of the same pedigree are likely to have the same genetic etiology, and extended pedigrees give more information about the genetic mechanisms underlying the trait than does a sample of unrelated persons. An important feature of pedigree data is that the statistical dependence of members in the same pedigree needs to be incorporated into analyses to obtain unbiased association parameters *(154)* and accurate estimates of the variances of these parameter estimates. For a large number of independent pedigrees, the robust method of generalized estimating equations has been used *(155)*. However, for one or a few pedigrees, the asymptotic results for

generalized estimating equations are not likely to hold, so it is necessary to use a statistical model that incorporates familial correlations. For continuous traits, methods based on the multivariate normal distribution with covariance matrices determined by the genetic relationships among pedigree members can be used *(156,157)*. Note, however, that this latter method and methods based on generalized estimating equations measure the marginal association of a genetic marker with a trait, not the association due to transmission/disequilibrium. Methods based on marginal associations are sensitive to population stratification and other nongenetic associations.

Alternative methods based on full likelihoods, which combine association, segregation, and linkage, have been proposed for nuclear families *(158,159)*; LD was incorporated by introducing coupling frequencies—the probabilities for a gamete to have the disease allele, given it carries a particular marker allele. One advantage of these likelihood models is that ill-defined ascertainment methods can be accounted for by conditioning on the affection status of everyone in the pedigree. Similar methods, called the MOD score, have been proposed for linkage analyses, by maximizing the LOD score over the genetic parameters (e.g., allele frequencies, penetrance, or recombination fraction). This method is a valid way to correct for complex and ill-defined ascertainment criteria *(34,35,160,161)* and is equivalent to maximizing the conditional probability of the marker genotypes, given the disease status, although there is some loss in efficiency. However, the model-based methods can be restrictive because the correct genetic model must be specified, and there is a large number of parameters to be estimated (e.g., allele frequencies, coupling frequencies, etc.). To avoid the long computer processing time required to fit these likelihood models and to avoid convergence problems when estimating many parameters, the Marker Association Segregation Chi-squares (MASC) method has been proposed *(162)*, which is based on estimating the genetic parameters of coupling frequencies and genotype-specific penetrances by minimizing a sum of independent Pearson chi-squares. It is worthwhile to note that 1) the likelihood methods are not sensitive to the absolute penetrance values but rather to their relative values *(162)*; 2) for diseases with low penetrance, the marker information from unaffected sibs is minimal [in fact, recent results regarding inclusion of unaffected sibs in *TDT*-like tests for complex disease indicate that this can do more harm than good, because unaffected sibs can add excessive variation to the marker genotype distribution *(163)*]; and 3) phenotype information from parents may be of importance *(164)*. Also, it has been shown that extension of the MASC method to account for two candidate genes involved in complex diseases can lead to increased power, instead of studying each separately *(111)*, much like what has been observed in genetic linkage analyses *(81)*.

Pooled Analyses

The detection of susceptibility genes for complex traits often requires large sample sizes that require collaboration of multiple research organizations. Also, multiple replication studies for linkage analyses of complex traits often produce different results. For these reasons, it is reasonable to predict that the discovery of genes underlying complex traits will often require pooling of data from multiple sources. Complications involved in performing these types of meta-analyses involve enforcing consistent diagnostic criteria, allowing for varying levels of diagnosis confirmation, and accounting for ethnic backgrounds and marker allele frequencies that vary across the different studies. Another factor that can affect pooled analyses is a difference in ascertainment criteria, which can cause the amount of linkage information to dramatically differ across different studies *(15)*. A popular meta-analytic method is to assume that the main parameters of interest are random variables that vary across the different studies. For the Haseman and Elston sib-pair linkage method, the regression parameter (for the regression of the squared trait difference on the estimated proportion of alleles IBD) can be assumed to depend on a random effect, so that the combined linkage result over all studies is a weighted average, giving more stable results than any single study can produce *(165)*. Another parameter of interest for parametric LOD score analyses is the fraction of linked families (the α parameter for the admixture test for linkage). However, this parameter should be interpreted with caution. First, the precision of the estimated α is typically very poor, resulting in a very wide confidence interval for α. Second, the estimated value of α can be biased when the assumed penetrance model is wrong. This is because, when the penetrance model is wrong, the recombination fraction is overestimated and because α and the recombination fraction are highly correlated, bias in the recombination fraction will cause bias in α. Robust methods for estimating α are needed for pooled analyses.

Technology Trends

The methodologic approaches applied to gene discovery are dictated, in part, by what is technically feasible within the laboratory. This constraint is currently in dramatic flux. Developing technologies may permit the realization of the promise of the routine application of genomic approaches to genetic analysis.

The genomic approach assesses the contribution of any gene or gene product, from the complete catalogue of genes and gene products, in defining a particular biologic state. Implicit in this approach is prior knowledge of the complete genome, or at least the genes that compose the genome. For this approach to be feasible, substantial information infrastructure and cost-effective comprehensive laboratory technologies are required.

Key to the success of these technologies are advances that have enhanced capabilities through increased miniaturization, parallelism, and automation. Miniaturization is required to reduce the laboratory reagent costs and sample volumes required to perform the large number of assays for comprehensive analysis. Miniaturization also offers formats that enable new applications. Parallelism is required to reduce the time necessary to complete the characterizations. For example, it is conservatively estimated that the human genome may contain 70 000 genes. To characterize one individual's genome in a year requires one to test 192 genes each day! Automation is required to address the large number of tests that must be performed, reduce costs, and improve quality control and uniformity of outcomes. On the basis of cost alone, it would not be feasible to perform large-scale genetic analysis without additional automation.

Technologies that are being proposed for genome-wide DNA analysis include serial sequencing, hybridization, and enzymatic approaches. In particular, hybridization array approaches have been developed in many formats to enable a wide range

of genome-wide DNA analyses (Table 1) *(166)*. Different technologies use very different types of capture sequences. Strategies for sequence capture include oligonucleotides (of varying length for different approaches) *(167)*, complementary DNAs (cDNAs) (either full length or fragments) *(168)*, and genomic DNA fragments. The method by which arrays are generated represents one of the more significant differences between technologies. The method associated with DNA "chips" uses the same photolithography technology used to make computer chips (the origin of the term "DNA chip") to build oligonucleotide detection sequences on a detection array *(169)*. An alternative strategy for building oligonucleotide sequences on glass slides uses technology similar to that used in ink jet printers. Other technologies simply "spot" oligos, cDNAs, or genomic DNA fragments at high density. Technologies also differ according to how sequences are discriminated and the capability to resolve single-base sequence variation. Different strategies include hybridization specificity, enzymatic specificity, and capture of solution-phase products. Fig. 3 shows an example of the microarray technology. In this example, hybridization to a cDNA microarray compares messenger RNA (mRNA) from one lymphoma specimen to a control mRNA pool from several lymphoma cell lines. Red spots are genes more highly expressed in the lymphoma specimen, and green spots are genes that are underexpressed in the lymphoma. The intensity of the spots can be used to indicate the magnitude of expression.

Table 1. Transcript expression analysis approaches and technology providers*

Sequencing of Expressed Sequence Tags (ESTs)
GeneTrace Systems
The Howard Hughes Medical Institute
IMAGE Consortium
Incyte
The Institute for Genomic Research
Serial Analysis of Gene Expression (SAGE)
Genzyme Molecular Oncology
The Johns Hopkins University, Drs. Kenneth Kinzler and Burt Vogelstein
Differential Display Approaches
Curagen Corporation, Quantitative Expression Analysis (QEA)
Digital Gene Technologies, Inc., Total Gene expression Analysis (TOGA)
Genaissance
GeneLogic, Restriction Enzyme Analysis of Differentially-Expressed Sequences (READS)
Hybridization Analysis
Affymetrix
Clontech
General Scanning, Inc.
Genetic MicroSystems, Inc.
GeneTrace Systems
Genome Systems
Genometrix
Hyseq, Inc.
Hyseq/Applied Biosystems Division of Perkin-Elmer
Incyte
Molecular Dynamics
National Human Genome Research Institute Laboratory of Cancer Genetics
Protogene
Research Genetics, Inc.
Stanford University, Dr. Pat Brown
Synteni
University of Washington, Dr. Lee Hood

*For a hypertext version of this table with more information and links to the individual WWW sites, go to http://www.ncbi.nlm.nih.gov/ncicgap/expression_tech_info.html. The information provided here is not comprehensive and is not meant to reflect an endorsement on the part of the National Cancer Institute of either the developers or the technologies.

The developing portfolio of genome-wide analysis technologies has many potential applications. Hybridization array technology has provided many examples of proof of concept experiments in genome-wide analysis. At the DNA level, arrays have been used for mutation detection within genes of interest, gene amplification or deletion detection, resequencing, and genotyping *(170–176)*. The array technology also has been demonstrated to be of utility in determining gene expression profiles for different cell types as well as in analysis of protein interactions and protein–DNA interactions. A similar breadth of applications is evolving for strategies using serial sequence and enzymatic approaches.

It is worth noting that few, if any, of these promising technologies have moved from proof-of-concept experiments conducted by the developer to routine application in other laboratories. Thus, the potential of comprehensive molecular analysis tools largely remains a promise. There are several reasons for this. First, much of the technology is insufficiently robust. In its current form, it is not suitable for day-to-day production application. The results, even in the hands of the developer, vary from experiment to experiment. Work is still required to "harden" the technology so that consistent results are routinely obtained. Another major obstacle to deployment of this new technology is its cost. Most of the approaches require substantial instrumentation and costly consumables. The cost of instrumentation for this technology is beyond the means of most academic investigators. Many of the reagents used in these assays are unique and must be custom generated. Therefore, the cost of development and manufacture must be recovered solely from the consumer, such as the research community (*see below*). A corollary of the cost obstacle is accessibility. Cost represents only one barrier to accessibility. Additional barriers include intellectual property issues, concern over market opportunities beyond the research community, and technology transfer logistics.

The rate at which this new technology is production "hardened" is determined only in a small part by the research community. The research community represents a very small market for many manufacturers and only recently have systems been developed that primarily target the basic and clinical research communities. To be cost-effective, instruments and reagents need to be produced on a very large scale and sold at prices that recoup development costs. Therefore, market forces beyond the needs of basic research drive the development and marketing of technology. Much of the current targeting of these efforts is to higher-end commercial users that require genomic information for the development of new therapeutics. Other potentially large markets include the areas of diagnostics, early detection (screening), and prevention; these areas remain underserved while awaiting technology hardening, reductions in cost, demonstration of utility, and an expanded base of information with respect to the function of expression products in the genome.

One of the greatest obstacles facing these new technologies is the conversion of data to information. Each of the new approaches promises to overwhelm the current information infrastructure. The challenge is not simply managing the tsunami of data that will be generated by these approaches. New analytic strategies will be required to interpret it and new models will need to be generated to allow investigators to manipulate the data. Each of the new approaches promises to overwhelm the

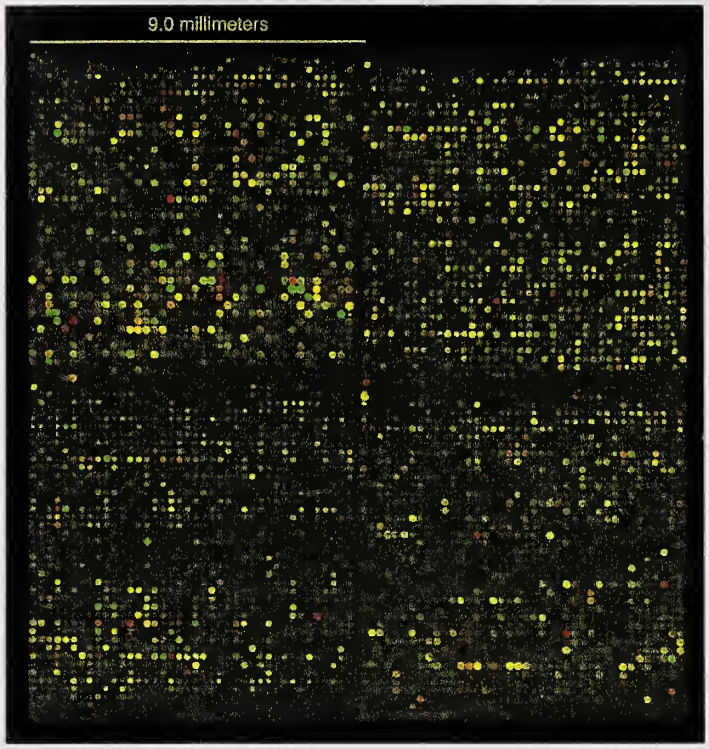


Fig. 3. Sample microarray image that represents the comparison of messenger RNA (mRNA) from one lymphoma specimen to a control mRNA pool from several lymphoma cell lines. **Red spots** are genes more highly expressed in the lymphoma specimen, and **green spots** are genes that are underexpressed in the lymphoma. [Image provided courtesy of A. Alizadeh and L. M. Staudt (Metabolism Branch, National Cancer Institute) and M. Eisen and P. Brown (Stanford University School of Medicine)].

current information infrastructure. The collection of interrelated challenges is illustrated in Fig. 4. With improved information analysis and modeling strategies, the promise exists for *in silico* modeling of biologic systems that can, at a minimum, generate hypotheses for testing as well as potentially predict the behavior of biologic systems.

The Changing Gene Discovery Paradigm

The introduction of genome-wide genetic analysis technologies has the capacity to dramatically alter the gene discovery paradigm. Gene discovery is currently dominated by the positional cloning paradigm. Its success has driven the generation of biologic reagents and analytic approaches. Genetic analysis is applied first to circumscribe the physical region of the genome that must be examined. Further genetic analysis is then used once genes within the region have been identified.

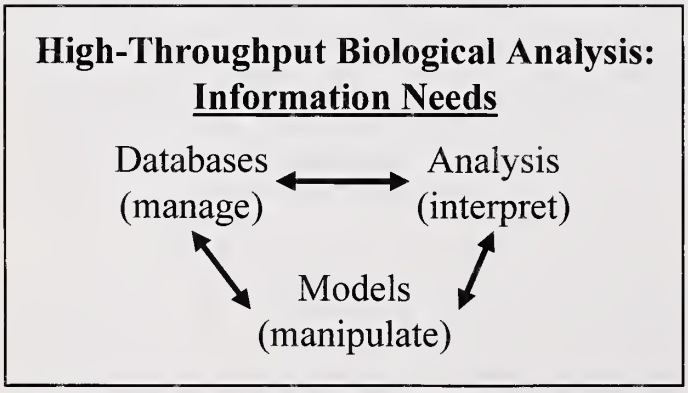


Fig. 4. Interrelationship of informatics needs associated with high-throughput genetic analysis. Only through the seamless integration of databases, models, and analysis tools will it be possible to extract information from the data generated by such large-scale analysis infrastructures.

However, one can argue that the current technologies, such as DNA chips and single nucleotide polymorphisms (SNPs) only make genotyping faster, cheaper, and more reliable, yet they do not seem to have changed fundamentally our basic concept of gene mapping, i.e., localization of the genes based on meiotic events. It is likely that the use of genome-wide analysis technology will alter this paradigm in the future. Gene discovery is currently dominated by the positional cloning paradigm, but an alternative approach starts with the genome's complement of genes (Fig. 5).

There is much enthusiasm for the genome-wide association studies *(177)*. In the simplest version of these studies, linkage is simply replaced by association as the metric for entering the physical mapping domain. Given that the physical regions that will be identified in these studies are smaller, the follow-up gene studies are likely to be much more manageable. It has been suggested that such studies can be performed with marker densities ranging from 3000 to 30 000 loci, depending on the evolutionary history of the population being studied. Developing microarray technology shows promise for permitting the routine characterization of this number of loci. However, it is worth noting that such high-resolution maps do not yet exist, and, even if they did exist, the hardened technologies for their routine characterization do not yet exist.

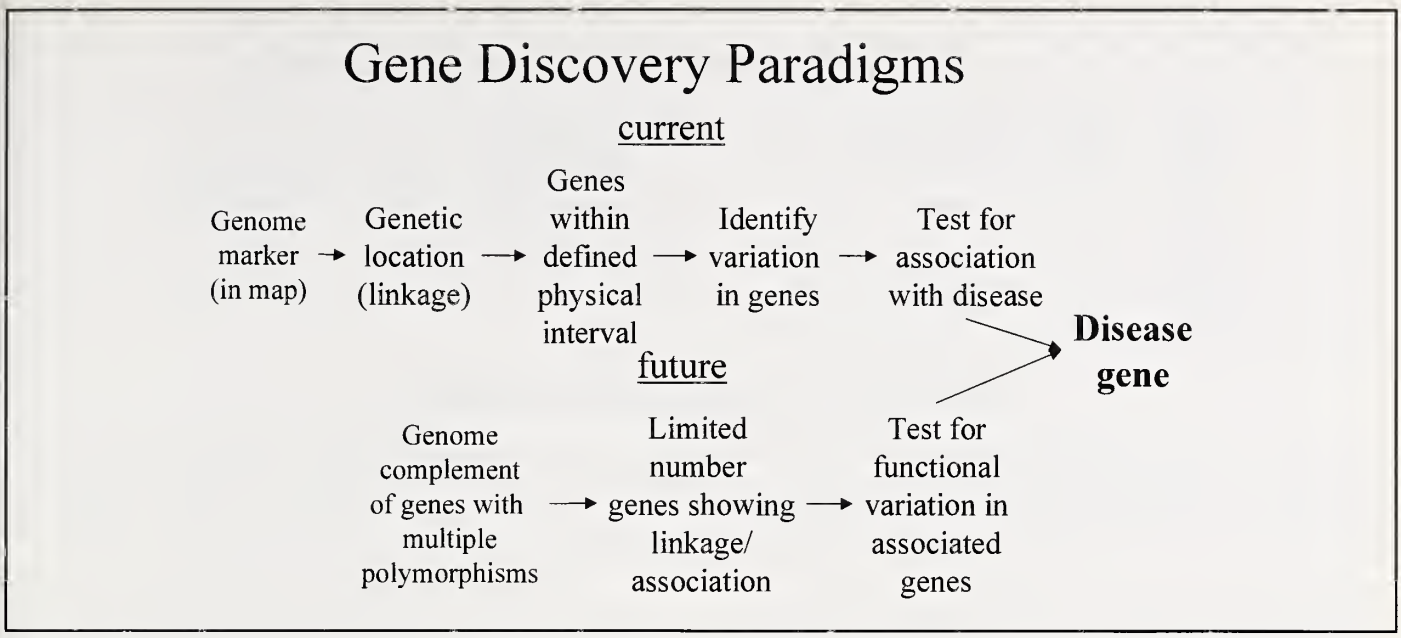


Fig. 5. Illustration of how the analytic gene discovery paradigms may change with the introduction of whole-genome technologies. The **upper panel** shows the steps in the current positional cloning paradigm. The **lower panel** indicates how discovery might change with the capacity to analyze all of the genes within the genome.

An alternative approach starts with the genome's complement of genes. Assume that each gene has multiple polymorphic "tags" in or near it. These tags could then be tested for association with disease status using a variety of study designs (*see above*). Those genes showing association are then candidates for more detailed functional and genetic analysis. The polymorphic marker density for this approach is within the same order of magnitude required for anonymous genome-wide disequilibrium mapping. Again, neither the reagents nor the technology exists today to permit this type of investigation. Of interest, the entire complement of genes within the human genome has yet to be definitively determined. Efforts such as the National Cancer Institute's Cancer Genome Anatomy Project *(178)* should address this critical limitation in the very near term.

It is possible to explore the potential of this approach with today's technology and resources. One approach is to use genetic analysis to assess the role of genes in well-described pathways in the development of disease. This approach merges gene mapping and "traditional" candidate locus studies by including as candidates all the members of a pathway. For complex traits, it is commonly not possible to construct a definitive set of candidate genes for study. This is because which portions of given pathways and which gene family members operate in different tissues following different exogenous or endogenous exposures is commonly unknown. When cost-effective genome technologies are available, it will be possible to routinely use genetic analysis to answer these questions.

With the use of existing technology, it has been possible to explore this approach for aflatoxin B_1 (AFB_1) exposure and primary hepatocellular carcinoma (HCC) *(179)*. AFB_1 is activated by the cytochrome P450 systems and detoxified by the glutathione *S*-transferase and epoxide hydrolase gene families. It is not *a priori* clear which members of these families should be most important in their respective roles. By tagging individual members with polymorphic variants and examining them in a case–control study *(180,181)*, it is possible to begin to assess which members may be important in modulating HCC risk through this pathway, as illustrated in Fig. 6.

Genetic Analysis Opportunities Novel to Cancer

The genetic analysis of cancer has available to it a resource not accessible to other complex traits—the tumor. Cancer is increasingly recognized as a disease of genes. It, therefore, stands to reason that information related to the inherited genetic basis of cancer may be obtained from genetic analysis of tumors. The tumor suppressor gene paradigm represents a historic example of the utility of joining somatic and germline genetic analysis *(182)*. Analytic methods through which somatic allele loss associated with tumor suppressor gene etiologies of cancer can be incorporated into linkage analysis have been previously described *(183,184)*.

Allele loss to localize tumor suppressor genes represents only one possibility for the use of somatic information in gene discovery. Comprehensive analysis of tumors has the potential of identifying molecular heterogeneity that may not be easily detectable by histopathology. This heterogeneity could be important in describing different etiologic pathways, both at a genetic level and at an environmental level. Different somatic patterns may provide clues to what exposures were important to cause a cancer and, therefore, clues into what genes may have mediated the exposures' effects.

Routine comprehensive somatic genetic characterization awaits the same reagent and technology development limiting the analyses described above. However, the potential utility of such information can be assessed in a limited way using today's technology and reagents. A collection of 32 HCC normal pairs was examined by the use of a collection of 391 genome-wide simple tandem repeat polymorphism markers. Allele loss patterns were observed to be complex. To assess whether underlying patterns existed within the data, evolutionary tree building algorithms (PHYLIP v3.5, provided courtesy of J. Felsenstein,

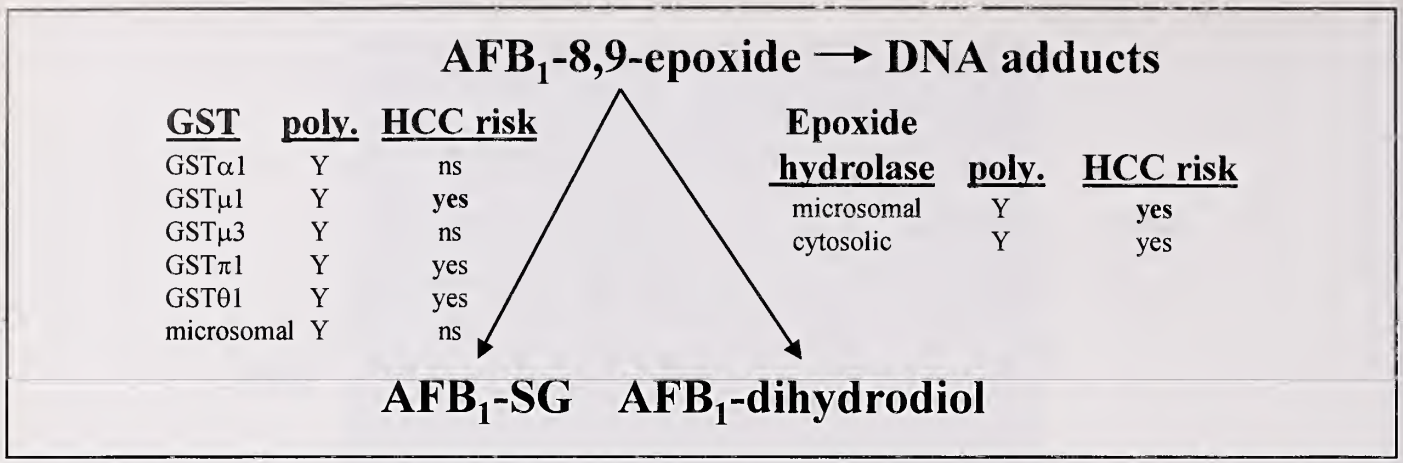


Fig. 6. Results of evaluating multiple genes in the aflatoxin B_1 (AFB_1) detoxification pathways. Multiple members of the glutathione *S*-transferase (GST) family and epoxide hydrolase families show allelic association with hepatocellular carcinoma (HCC). "Yes" in association column indicates a significant association; "ns" means no significant association observed. Poly = polymorphic and SG = glutathione conjugate.

http://evolution.genetics.washington.edu/phylid.html) were used to examine the data. Markers informative in more than 10 tumors were analyzed by MIX and CONSENSE. CONSENSE identified a tree consistent at the first set of nodes (found in 100% of replicates). The branches of the consensus tree (Fig. 7) had different chromosome-specific loss patterns, different rates of genome-wide allele loss rates, and different candidate locus risk allele distributions. This finding suggests that inclusion of such information may be important in gene discovery exercises.

Summary

In summary, we have reviewed a broad range of design and analytic issues related to the discovery of cancer susceptibility

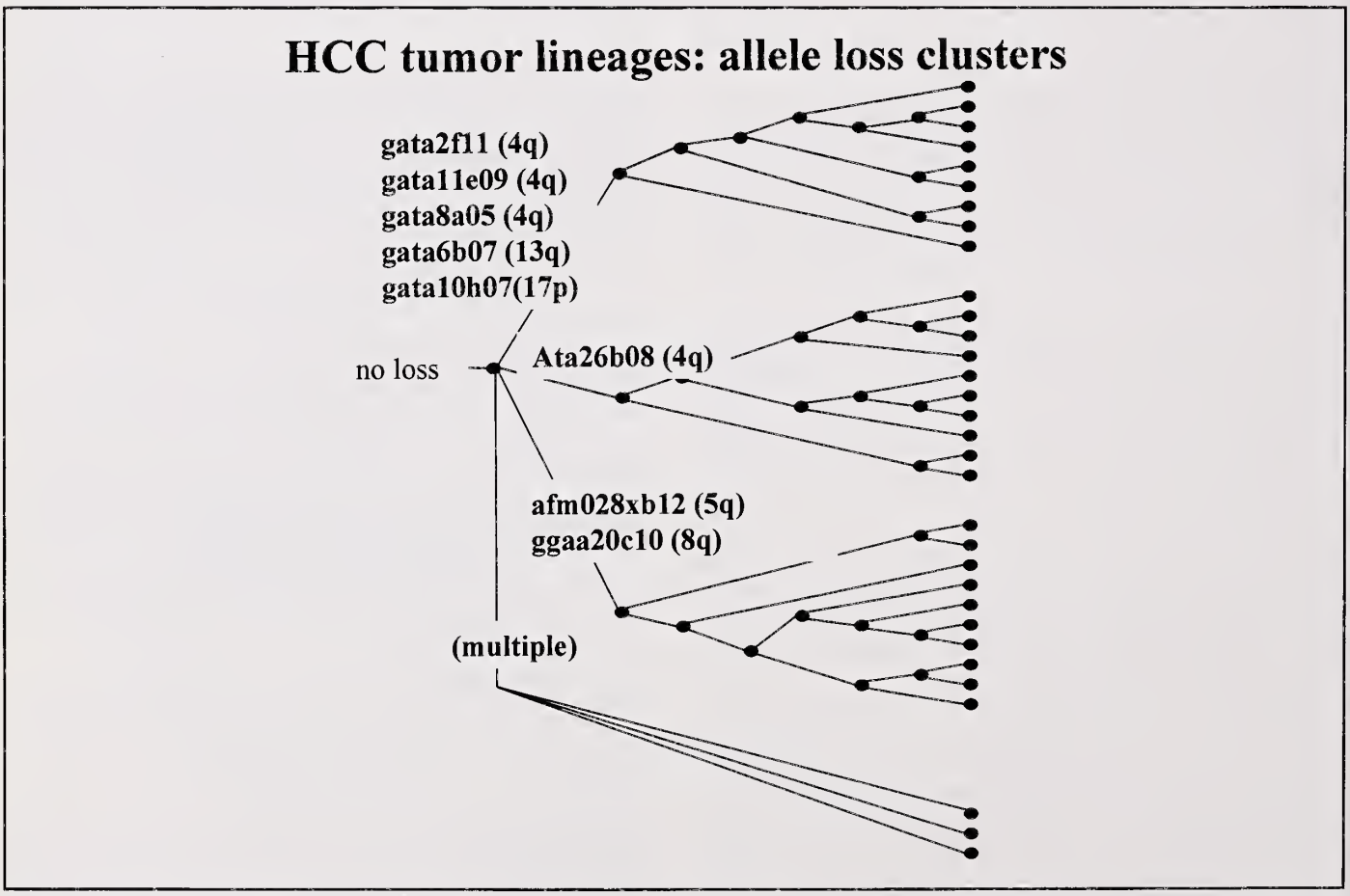


Fig. 7. Tree constructed using allele loss data showing clusters of related tumors. Loci indicated at branch points indicate genomic locations showing allele loss in all tumors within that cluster. Resampling analysis indicates that 100% of replicates show this pattern of loss. HCC = hepatocellular carcinoma.

genes. Some critical points considered were the following: the definition of phenotype and the necessary efforts required to verify diagnoses; sampling issues regarding ASPs, nuclear families, and large extended pedigrees; analytic methods for gene discovery, including linkage strategies, and association methods; and new genetic technology for genomic analyses. To fully exploit genetic information and to improve efficiency of gene discovery, it will be necessary to have adequate samples available, such as by well-defined family-based registries, as well as continued development of improved study designs, genomic technology, and analytic methods.

References

(1) Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 1990;46:222–8.

(2) Lindor NM, Greene MH, the Mayo Familial Cancer Program. The concise handbook of family cancer syndromes. J Natl Cancer Inst 1998;14:1039–71.

(3) Airewele G, Adatto P, Cunningham J, Mastromarino C, Spencer C, Sharp M. Family history of cancer in patients with glioma: a validation study of accuracy. J Natl Cancer Inst 1998;90:543–4.

(4) Love RR, Evans AM, Josten DM. The accuracy of patient reports of a family history of cancer. J Chron Dis 1985;38:289–93.

(5) Pepe M, Fleming TR. A nonparametric method for dealing with mismeasured covariate data. J Am Stat Assoc 1991;86:108–13.

(6) Wijsman EM, Amos CI. Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. Genet Epidemiol 1997;14:719–35.

(7) Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8:431–40.

(8) Pepe MS. Inference using surrogate outcome data and a validation sample. Biometrika 1992;79:355–65.

(9) Ginsburg EK, Axenovich TI. On planning of samples for linkage analysis: two ways of a sample size reduction. Genet Epidemiol 1996;13:343–54.

(10) Risch N. Mapping genes for psychiatric disorders. In: Gershon ES, Cloninger CR, editors. Genetic approaches to mental disorders. Washington (DC): American Psychiatric Press; 1994. p. 47–61.

(11) Weeks DE. Pedigree selection and information content. Curr Protocols Hum Genet 1994;1:1–21.

(12) Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 1987;236:1567–70.

(13) Pauls DL. Behavioural disorders: lessons in linkage. Nat Genet 1993;3:4–5.

(14) Greenberg DA. There is more than one way to collect data for linkage analysis: what a study of epilepsy can tell us about linkage strategy for psychiatric disease. Arch Gen Psychiatry 1992;49:745–50.

(15) McCarthy MI, Kruglyak L, Lander ES. Sib-pair collection strategies for complex diseases. Genet Epidemiol 1998;15:317–40.

(16) Goldin LR, Bailey-Wilson JE, Borecki IB, Falk CT, Goldstein AM, Suarez BK, et al. Genetic analysis workshop 10: detection of genes for complex traits. Genet Epidemiol 1997;14:549–1152.

(17) Goldgar DE, Easton DF. Optimal strategies for mapping complex diseases in the presence of multiple loci. Am J Hum Genet 1997;60:1222–32.

(18) Schork NJ, Xu X. Sib pairs versus pedigrees: what are the advantages? Diab Rev 1997;5:116–22.

(19) Ploughman LM, Boehnke M. Estimating the power of a proposed linkage study for a complex trait. Am J Hum Genet 1989;44:543–51.

(20) Ginsburg EK, Axenovich TI, Goodman DW. On estimation of linkage test power. Genet Epidemiol 1996;13:355–65.

(21) Ginsburg EK, Axenovich TI. Sample size required for predefined linkage decision quality. Genet Epidemiol 1997;14:479–91.

(22) Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 1990;46:229–41.

(23) Todorov AA, Borecki IB, Rao DC. Linkage analysis of complex traits using affected sibpairs: effects of single-locus approximations on estimates of the required sample size. Genet Epidemiol 1997;14:389–401.

(24) Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science 1995;268:1584–9.

(25) Zhao H, Zhang H, Rotter JI. Cost-effective sib-pair designs in the mapping of quantitative-trait loci. Am J Hum Genet 1997;60:1211–21.

(26) Field LL, Tobias R, Magnus T. A locus on chromosome 15q26 (IDDM3) produces susceptibility to insulin-dependent diabetes mellitus. Nat Genet 1994;8:189–94.

(27) Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, et al. A genome-wide search for human type 1 diabetes susceptibility genes. Nature 1994;371:130–6.

(28) Hashimoto L, Habita C, Beressi JP, Delepine M, Besse C, Cambon-Thomsen A, et al. Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 1994;371:161–4.

(29) Concannon P, Gogolin-Ewens KJ, Hinds DA, Walpelhorst B, Morrison VA, Stirling B, et al. A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. Nat Genet 1998;19:292–6.

(30) Mein CA, Esposito L, Dunn MG, Johnson GC, Timms AE, Goy JV, et al. A search for type 1 diabetes susceptibility genes in families from the United Kingdom. Nat Genet 1998;19:297–300.

(31) Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7.

(32) Lander ES. The new genomics: global views of biology. Science 1996;274:536–9.

(33) Elston RC. The genetic dissection of multifactorial traits. Clin Exp Allergy 1995;25:103–6.

(34) Clerget-Darpoux F, Bonaiti-Pellie C. Strategies based on marker information for the study of human diseases. Ann Hum Genet 1992;56:145–53.

(35) Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J. Effects of misspecifying genetic parameters in lod score analysis. Biometrics 1986;42:393–9.

(36) Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994;265:2037–48.

(37) Clerget-Darpoux F. Bias of the estimated recombination fraction and lod score due to an association between a disease gene and a marker gene. Ann Hum Genet 1982;46:363–72.

(38) Manuila A. Blood groups and disease—hard facts and delusions. JAMA 1958;167:2047–53.

(39) Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet 1955;7:277–318.

(40) Whittemore AS. Genome scanning for linkage: an overview. Am J Hum Genet 1996;59:704–16.

(41) Ott J. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am J Hum Genet 1974;26:588–97.

(42) Lathrop GM, Lalouel JM. Easy calculation of lod scores and genetic risks on small computers. Am J Hum Genet 1984;36:460–5.

(43) Cottingham RW Jr, Idury RM, Schaffer AA. Faster sequential genetic linkage computations. Am J Hum Genet 1993;53:252–63.

(44) Schaffer AA, Gupta SK, Shiram K, Cottingham RW Jr. Avoiding recomputation in linkage analysis. Hum Hered 1994;44:225–37.

(45) O'Connell JR, Weeks DE. The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-redcoding and fuzzy inheritance. Nat Genet 1995;11:402–8.

(46) Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 1996;58:1347–63.

(47) Amos CI, Williamson JA. Robustness of the maximum-likelihood (LOD) method for detecting linkage. Am J Hum Genet 1993;52:213–4.

(48) Williamson JA, Amos CI. Guess LOD approach: sufficient conditions for robustness. Genet Epidemiol 1995;12:163–76.

(49) Weeks DE, Lange K. The affected-pedigree-member method of linkage analysis. Am J Med Genet 1988;42:315–26.

(50) Davis S, Weeks DE. Comparison of nonparametric statistics for detection of linkage in nuclear families: single-marker evaluation. Am J Hum Genet 1997;61:1431–44.

(51) Weeks DE, Lathrop GM. Polygenic disease: methods for mapping complex disease traits. Trends Genet 1998;11:513–9.

(52) Whittemore AS, Tu IP. Simple, robust linkage tests for affected sib pairs. Am J Hum Genet 1998;62:1228–42.

(53) Sandkuyl LA. Analysis of affected sib pairs using information from extended families. Prog Clin Bio Res 1989;329:117–22.

(54) Sham PC, Zhao JH, Curtis D. Optimal weighting scheme for affected sib-pair analysis of sibship data. Ann Hum Genet 1997;61:61–9.
(55) Curtis D, Sham PC. Using risk calculation to implement an extended relative pair analysis. Ann Hum Genet 1994;58:151–62.
(56) Hauser ER, Boehnke M, Guo SW, Risch N. Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. Genet Epidemiol 1996;13:117–37.
(57) Holmans P. Asymptotic properties of affected-sib-pair linkage analysis. Am J Hum Genet 1993;52:362–74.
(58) Knapp M, Seuchter SA, Baur MP. Two-locus disease models with two marker loci: the power of affected-sib-pair tests. Am J Hum Genet 1994; 55:1030–41.
(59) Cordell HJ, Todd JA, Bennett ST, Kawaguchi Y, Farrall M. Two-locus maximum lod score analysis of a multifactorial trait: joint consideration of IDDM2 and IDDM4 with IDDM1 in type 1 diabetes. Am J Hum Genet 1995;57:920–34.
(60) Farrall M. Affected sibpair linkage tests for multiple linked susceptibility genes. Genet Epidemiol 1997;14:103–15.
(61) Whittemore AS, Halpern J. A class of tests for linkage using affected pedigree members. Biometrics 1994;50:118–27.
(62) Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 1997;61:1179–88.
(63) Knapp M, Seuchter SA, Baur MP. Linkage analysis in nuclear families 1: optimality criteria for affected sib-pair tests. Hum Hered 1994;44:37–43.
(64) Schroeder M, Brown DL, Weeks DE. Improved programs for the affected-pedigree-member method of linkage analysis. Genet Epidemiol 1994;11:69–74.
(65) Ward PJ. Some developments on the affected-pedigree method of linkage analysis. Am J Hum Genet 1993;52:1200–15.
(66) Weeks DE, Valappil TI, Schroeder M, Brown DL. An X-linked version of the affected-pedigree-member method of linkage analysis. Hum Hered 1995;45:25–33.
(67) Weeks DE, Lange K. A multilocus extension of the affected-pedigree-member method of linkage analysis. Am J Hum Genet 1992;50:859–68.
(68) Matise TC, Weeks DE. Detecting heterogeneity with the affected-pedigree-member (APM) method. Genet Epidemiol 1993;10:401–6.
(69) Weeks DE, Harby LD. The affected-pedigree-member method: power to detect linkage. Hum Hered 1995;45:13–24.
(70) Bishop DT, Williamson JA. The power of identity-by-state methods in linkage analysis. Am J Hum Genet 1990;46:254–65.
(71) Goldin LR, Weeks DE. Two-locus models of disease: comparison of likelihood and nonparametric linkage methods. Am J Hum Genet 1993; 53:908–15.
(72) Davis S, Schroeder M, Goldin LR, Weeks D. Nonparametric simulation-based statistics for detecting linkage in general pedigrees. Am J Hum Genet 1996;58:867–80.
(73) Babron MC, Martinez M, Bonaiti-Pellie C, Clerget-Darpoux F. Linkage detection by the affected-pedigree-member method: what is really tested? Genet Epidemiol 1993;10:389–94.
(74) Commenges D, Olson J, Wijsman E. The weighted rank pairwise correlation statistic for linkage analysis: simulation study and application to Alzheimer's disease. Genet Epidemiol 1994;11:201–12.
(75) Commenges D. Robust genetic linkage analysis based on a score test of homogeneity: the weighted pairwise correlation statistic. Genet Epidemiol 1994;11:189–200.
(76) Commenges D, Laurent A. Improving the robustness of the weighted pairwise correlation test for linkage analysis. Genet Epidemiol 1996;13: 559–73.
(77) Greenberg DA, Hodge SE, Vieland VJ, Spence MA. Affecteds-only linkage methods are not a panacea. Am J Hum Genet 1996;58:892–5.
(78) Farrall M. LOD wars: the affected-sib-pair paradigm strikes back. Am J Hum Genet 1997;60:735–7.
(79) Kruglyak L. Nonparametric linkage tests are model free. Am J Hum Genet 1997;61:254–5.
(80) Greenberg DA, Hodge SE, Vieland VJ, Spence MA. Power, model of inheritance, and Type I error in lod scores and affecteds-only methods: reply to Kruglyak. Am J Hum Genet 1998;62:202–4.
(81) Schork NJ, Boehnke M, Terwilliger JD, Ott J. Two-trait-locus linkage analysis: a powerful strategy for mapping complex genetic traits. Am J Hum Genet 1993;53:1127–36.
(82) Martinez M, Goldin LR. Power of the linkage test for a heterogeneous disorder due to two independent inherited causes: a simulation study. Genet Epidemiol 1990;7:219–30.
(83) Levinson DF. Power to detect linkage with heterogeneity in samples of small nuclear families. Am J Med Genet 1993;48:94–102.
(84) Goldin LR, Gershon ES. Power of the affected-sib-pair method for heterogeneous disorders. Genet Epidemiol 1988;5:35–42.
(85) Bishop CM. Neural networks for pattern recognition. Oxford (U.K.): Clarendon Press; 1995.
(86) Boehnke M. Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. Am J Hum Genet 1994;55:379–90.
(87) Jorde LB. Invited editorial: linkage disequilibrium as a gene-mapping tool. Am J Hum Genet 1995;56:11–4.
(88) Devlin B, Risch N, Roeder K. Disequilibrium mapping: composite likelihood for pairwise disequilibrium. Genomics 1996;36:1–16.
(89) Rannala B, Slatkin M. Likelihood analysis of disequilibrium mapping, and related problems. Am J Hum Genet 1998;62:459–73.
(90) Lazzeroni LC. Linkage disequilibrium and gene mapping: an empirical least-squares approach. Am J Hum Genet 1998;62:159–70.
(91) Guo SW. Linkage disequilibrium measures for fine-scale mapping: a comparison. Hum Hered 1997;47:301–14.
(92) Xiong M, Guo SW. Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. Am J Hum Genet 1997;60:1513–31.
(93) Terwilliger JD. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am J Hum Genet 1995;56:777–87.
(94) Hastbacka J, delaChapelle A, Kaitila I, Sistonen P, Weaver A, Lander E. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat Genet 1992;2:204–11.
(95) Kaplan NL, Hill WG, Weir BS. Likelihood methods for locating disease genes in nonequilibrium populations. Am J Hum Genet 1995;56:18–32.
(96) Boguski MS. The turning point in genome research. Trends Biochem Sci 1995;20:295–6.
(97) Walter MA, Spillett DJ, Thomas P, Weissenbach J, Goodfellow PN. A method for constructing radiation hybrid maps of whole genomes. Nat Genet 1994;7:22–8.
(98) Houwen RH, Baharloo S, Blankenship K, Raeymakers P, Juyn F, Sandkuijl LA, et al. Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet 1994;8:380–6.
(99) Stephens JC, Briscoe D, O'Brien SJ. Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am J Hum Genet 1994;55:809–24.
(100) Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci U S A 1988;85:9119–23.
(101) Kaplan NL, Martin ER, Morris RW, Weir BS. Marker selection for the transmission/disequilibrium test, in recently admixed populations. Am J Hum Genet 1998;62:703–12.
(102) Chapman NH, Wijsman EM. Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. Am J Hum Genet 1998;63:1872–85.
(103) Xiong M, Guo SW. The power of linkage detection by the transmission/disequilibrium tests. Hum Hered 1998;48:295–312.
(104) Falk CT, Rubinstein P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 1987;51:227–33.
(105) Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506–16.
(106) Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 1996;59:983–9.
(107) Terwilliger JD, Ott J. A haplotype-based 'Haplotype Relative Risk' approach to detecting allelic associations. Hum Hered 1992;42:337–46.
(108) Knapp M, Seuchter SA, Bauer MP. The haplotype-relative-risk (HRR) method for analysis of association in nuclear families. Am J Hum Genet 1993;52:1085–93.
(109) Ott J. Statistical properties of the haplotype relative risk. Genet Epidemiol 1989;6:127–30.

(110) Parsian A, Todd RD, Devor EJ, O'Malley KL, Suarez, BK, Reich T, et al. Alcoholism and alleles of the human D2 dopamine receptor locus. Studies of association and linkage. Arch Gen Psychiatry 1991;48: 655–63.
(111) Schaid DJ, Sommer SS. Genotype relative risks: methods for design and analysis of candidate-gene association studies. Am J Hum Genet 1993; 53:1114–26.
(112) Schaid DJ, Sommer SS. Comparison of statistics for candidate-gene association studies using cases and parents. Am J Hum Genet 1994;55: 402–9.
(113) Jin K, Speed TP, Klitz W, Thomson G. Testing for segregation distortion in the HLA complex. Biometrics 1994;50:1189–98.
(114) Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 1995;57:455–64.
(115) Thomson G. Analysis of complex human genetic traits: an ordered-notation method and new tests for mode of inheritance. Am J Hum Genet 1995;57:474–86.
(116) Thomson G. Mapping disease genes: family-based association studies. Am J Hum Genet 1995;57:487–98.
(117) Flanders WD, Khoury MJ. Analysis of case-parental control studies: method for the study of associations between genetic markers. Am J Epidemiol 1996;144:696–703.
(118) Khoury MJ. Case-parental control method in the search for disease-susceptibility genes. Am J Hum Genet 1994;55:410–5.
(119) Tai JJ, Song WH. Linkage disequilibrium and linkage information from one-child families. Hum Hered 1991;41:316–23.
(120) Harley JB, Moser KL, Neas BR. Logistic transmission modeling of simulated data. Genet Epidemiol 1995;12:607–12.
(121) Clerget-Darpoux F, Babron MC, Bickeboller H. Comparing the power of linkage detection by the transmission disequilibrium test and the identity-by-descent test. Genet Epidemiol 1995;12:583–8.
(122) Morris AP, Curnow RN, Whittaker JC. Randomization tests of disease-marker associations. Ann Hum Genet 1997;61:49–60.
(123) Rice JP, Neuman RJ, Hoshaw SL, Daw EW, Gu C. TDT with covariates and genomic screens with mod scores: their behavior on simulated data. Genet Epidemiol 1995;12:659–64.
(124) Sham PC, Curtis D. An extended transmission/equilibrium test (TDT) for multi-allele marker loci. Ann Hum Genet 1995;59:323–36.
(125) Bickeboller H, Clerget-Darpoux F. Statistical properties of the allelic and genotypic transmission/disequilibrium test for multiallelic markers. Genet Epidemiol 1995;12:865–70.
(126) Kaplan NL, Martin ER, Weir BS. Power studies for the transmission/disequilibrium tests with multiple alleles. Am J Hum Genet 1997;60: 691–702.
(127) Sham P. Transmission/disequilibrium tests for multiallelic loci. Am J Hum Genet 1997;61:774–8.
(128) Cleves MA, Olson JM, Jacobs KB. Exact transmission-disequilibrium tests with multiallelic markers. Genet Epidemiol 1997;14:337–47.
(129) McKeigue PM. Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. Am J Hum Genet 1997;60:188–96.
(130) Camp NJ. Genomewide transmission/disequilibrium testing—consideration of the genotypic relative risk at disease loci. Am J Hum Genet 1997;61:1424–30.
(131) Martin ER, Kaplan NL, Weir BS. Tests for linkage and association in nuclear families. Am J Hum Genet 1997;61:439–48.
(132) Morris AP, Whittaker JC, Curnow RN. A likelihood ratio test for detecting patterns of disease-marker association. Ann Hum Genet 1997;61: 335–50.
(133) Maestri NE, Beaty TH, Hetmanski J, Smith EA, McIntosh I, Wyszynski DF, et al. Application of transmission disequilibrium tests to nonsyndromic oral clefts: including candidate genes and environmental exposures in the models. Am J Med Genet 1997;73:337–44.
(134) Wilson SR. On extending the transmission/disequilibrium test (TDT). Ann Hum Genet 1997;61:151–61.
(135) Self SG, Longton G, Kopecky KJ, Liang KY. On estimating HLA/disease association with application to a study of aplastic anemia. Biometrics 1991;47:53–61.
(136) Knapp M, Wassmer G, Baur MP. The relative efficiency of the Hardy-Weinberg equilibrium-likelihood and the conditional on parental genotype-likelihood methods for candidate-gene association studies. Am J Hum Genet 1995;57:1476–85.
(137) Schaid DJ. General score tests for associations of genetic markers with disease using cases and their parents. Genet Epidemiol 1996;13:423–49.
(138) Langholz B, Tuomilehto-Wolf E, Thomas D, Pitkaniemi J, Tuomilehto J, DiMe Study Group. Variation in HLA-associated risks of childhood insulin dependent diabetes in the Finnish population: I. Allele effects at A, B, and DR loci. Genet Epidemiol 1995;12:441–53.
(139) Schaid DJ. Relative-risk regression models using cases and their parents. Genet Epidemiol 1995;12:813–8.
(140) Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case–control studies of candidate genes and gene–environment interactions: basic family designs. Am J Epidemiol 1999;149:693–705.
(141) Goldstein AM, Hodge SE, Haile RW. Selection bias in case–control studies using relatives as the controls. Int J Epidemiol 1989;18:985–9.
(142) Andrieu N, Goldstein AM. Use of relatives of cases as controls to identify risk factors when an interaction between environmental and genetic factors exists. Int J Epidemiol 1996;25:649–55.
(143) Curtis D. Use of siblings as controls in case–control association studies. Ann Hum Genet 1997;61:319–33.
(144) Langefeld CD, Pericak-Vance MA, Saunders AM, Boehnke M. Family-based tests for association using discordant sib pairs. Am J Hum Genet 1997;61:A1643.
(145) Monks SA, Martin ER, Weir BS, Kaplan NL. A sibship test of linkage in the absence of parental information. Am J Hum Genet 1997;61:A1669.
(146) Ewens WJ, Spielman RS. The sib-TDT (S-TDT): a TDT (transmission/disequilibrium test) without parents. Am J Hum Genet 1997;61:A1600.
(147) Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 1998;62:450–8.
(148) Schaid DJ, Rowland CR. Parents, sibs, and unrelated controls to detect associations of genetic markers with disease. Am J Hum Genet 1998;63: 1492–1506.
(149) Long JC, Williams RC, Urbanek M. An E-M algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 1995;56:799–810.
(150) Excoffier L, Slatkin M. Incorporating genotypes of relatives into a test of linkage disequilibrium. Am J Hum Genet 1998;62:171–80.
(151) Schweder T, Spjotvoll E. Plots of *P*-values to evaluate many tests simultaneously. Biometrika 1982;69:493–502.
(152) Thomas D, Pitkaniemi J, Langholz B, Tuomilehto-Wolf E, Tuomilehto J. Variation of HLA-associated risks of childhood insulin-dependent diabetes in the Finnish population: II. Haplotype effects. Genet Epidemiol 1995;12:455–66.
(153) Thomas D, Langholz B, Clayton D, Pitkaniemi J, Tuomilehto-Wolf E, Toumilehto J. Emperical Bayes methods for testing associations with large numbers of candidate genes in the presence of environmental risk factors, with applications to HLA associations in IDDM. Ann Med 1992; 24:387–92.
(154) Liang KY. Extended Mantel–Haenszel estimating procedure for multivariate logistic regression models. Biometrics 1987;43:289–99.
(155) Tregouet DA, Ducimetiere P, Tiret L. Testing association between candidate-gene markers and phenotype in related individuals, by use of estimating equations. Am J Hum Genet 1997;61:189–99.
(156) Boerwinkle E, Chakraborty R, Sing CF. The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. Ann Hum Genet 1986;50:181–94.
(157) George VT, Elston RC. Testing the association between polymorphic markers and quantitative traits in pedigrees. Genet Epidemiol 1987;4: 193–201.
(158) Risch N. Segregation analysis incorporating linkage markers. 1. Single-locus models with an application to Type I diabetes. Am J Hum Genet 1984;36:363–86.
(159) Maclean CJ, Morton NE, Yee S. Combined analysis of genetic segregation and linkage under an oligogenic model. Comp Biomed Res 1984;17: 471–80.
(160) Elston RC. Man bites dog? The validity of maximizing lod scores to determine mode of inheritance. Am J Med Genet 1989;34:487–8.
(161) Hodge SE, Elston RC. Lods, wrods, and mods: the interpretation of lod scores calculated under different models. Genet Epidemiol 1994;11:329–42.

(162) Clerget-Darpoux F, Babron MC, Prum B, Lathrop GM, Deschamps I, Hors J. A new method to test genetic models in HLA associated diseases: the MASC method. Ann Hum Genet 1988;52:247–58.

(163) Boehnke M, Langefeld CD. A transmission/disequilibrium test that uses both affected and unaffected offspring. Am J Hum Genet 1997;61: A15464.

(164) Ewens WJ, Spielman RS. Statistical properties of maximum likelihood estimators for genetic parameters of HLA-linked diseases. Am J Hum Genet 1985;37:1172–91.

(165) Li Z, Rao DC. Random effects model for meta-analysis of multiple quantitative sibpair linkage studies. Genet Epidemiol 1996;13:377–83.

(166) Marshall A, Hodgson J. DNA chips: an array of possibilities. Nat Biotech 1998;16:27–31.

(167) Fodor SPA, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL. Multiplexed biochemical assays with biological chips. Nature 1993;364: 555–6.

(168) Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467–70.

(169) Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A 1994;91:5022–6.

(170) Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotech 1996;14:1675–80.

(171) DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, et al. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 1996;14:457–60.

(172) Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. Nat Genet 1996;14:441–7.

(173) Shoemaker DD, Lashkare DA, Morries D, Mittmann M, Davis RW. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat Genet 1996;14: 450–6.

(174) Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, et al. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc Natl Acad Sci U S A 1997;94:13057–62.

(175) Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 1998;280:1077–82.

(176) Drmanac S, Kita D, Labat K, Hauser B, Schmidt C, Burczak JD, et al. Accurate sequencing by hybridization for DNA diagnostics and individual genomics. Nat Biotech 1998;16:54–8.

(177) Collins FS, Guyer MS, Chakravarti A. Variations on a theme: cataloging human DNA sequence variation. Science 1998;278:1580–1.

(178) Strausberg RL, Dahl CA, Klausner RD. New opportunities for uncovering the molecular basis of cancer. Nat Genet 1997;16(suppl):415–516.

(179) McGlynn KA, Rosvold EA, Lustbader ED, Hu Y, Clapper ML, Zhou T, et al. Susceptibility to hepatocellular carcinoma is associated with genetic variation in the enzymatic detoxification of aflatoxin B1. Proc Natl Acad Sci U S A 1995;92:2384–7.

(180) Murray JC, Buetow KH, Donovan M, Hornung S, Motulsky AG, Disteche C, et al. Linkage disequilibrium of plasminogen polymorphisms and assignment of the gene to human chromosome 6q26–6q27. Am J Hum Genet 1987;40:338–50.

(181) Ardinger HH, Buetow KH, Bell Gl, Bardach J, VanDemark DR, Murray JC. Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. Am J Hum Genet 1989;45:348–53.

(182) Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68:820–3.

(183) Rebbeck TR, Lustbader ED, Buetow KH. Somatic allele loss in genetic linkage analysis of cancer. Genet Epidemiol 1994;11:419–29.

(184) Lustbader ED, Rebbeck TR, Buetow KH. Using loss of heterozygosity data in affected pedigree member linkage tests. Genet Epidemiol 1995; 12:339–50.

II. GENE CHARACTERIZATION PANEL

Design of Gene Characterization Studies: an Overview

Duncan C. Thomas

This collection of papers from the Gene Characterization Panel addresses design issues in studies aimed at assessing the population characteristics of cloned genes, such as their allele frequencies, penetrance, variation in these parameters across subpopulations, and gene–environment and gene–gene interactions. This paper provides an overview of the various designs that have been suggested, including cohort and case–control designs using independent and related individuals as well as optimal multistage sampling and hybrid designs. Various statistical (bias and efficiency) and practical considerations are suggested for evaluation of the alternative designs, with the aim of posing the question, "What is the optimal design for a particular situation"? The answer to this question clearly depends on such contextual issues as nature of the outcome variable, the gene frequency and genetic relative risk, and the importance of gene–environment and gene–gene interactions. Further methodologic work might be usefully directed toward assessment of the seriousness of the population stratification problem in general as well as methods of dealing with it, the utility of registries of high-risk families, and the merits of various hybrid designs for gene discovery and gene characterization. [Monogr Natl Cancer Inst 1999;26:17–23]

Aims of Characterization Studies

The class of "gene characterization studies" includes those studies involving a measured gene that is either known to be causally related to some trait or for which such an association is to be tested. Thus, studies of linkage or association with marker genes, or both, that are not thought to be causal factors but merely tools for mapping causal genes are the subject of the Gene Discovery Panel in this volume.

Before proceeding further, it is worth pausing to clarify the meaning of the word "causal" in the genetic context. A gene is a causal risk factor for a trait if the expected outcome depends on an individual's genotype at that locus, all other risk factors (genetic and environmental) being held constant. For a dichotomous disease trait, the expected outcome is the age-specific (and all-other-factors-specific) incidence rate, otherwise known as "penetrance." Thus, a gene is causal if this penetrance varies by genotype, after adjusting for all other causal factors. In principle, this definition would exclude associations with marker genes that are due entirely to population associations with other causal genes, but in practice the definition is hard to operationalize, since it might require adjusting for causal loci that have not yet been identified or that are in complete linkage disequilibrium with the truly causal locus. But at a minimum, a locus must be functionally significant to be considered causal.

In the early stages of an investigation, one might wish to test for an association with a candidate gene that has been suggested as a possible causal agent, perhaps by virtue of being located in a region suggested by linkage studies, perhaps by virtue of knowledge of some metabolic function that is plausibly related to the disease process, or perhaps by homology with a known causal gene in other species. In statistical terms, we are interested in testing the null hypothesis that the genetic relative risk associated with this gene is unity. Of course, rejection of the statistical null hypothesis does not by itself establish causality; such a judgment still depends on other criteria familiar to epidemiologists, such as absence of confounding (especially by other genes) or bias in the study design, biologic plausibility (e.g., a functional pathway), consistency across studies, and concordance with other types of knowledge.

Once a gene has been cloned and its causal connection with the disease has been reasonably well established, many questions remain about its quantitative aspects. These aspects can be broadly grouped into two categories: population characteristics and risk characteristics. Under the heading of population characteristics, we are primarily interested in the prevalence of alleles and the variation in prevalence with various population characteristics, such as ethnicity or other genes with which it is in disequilibrium. Under the heading of risk characteristics are all aspects of the dependence of penetrance on genotype. Foremost among these is the main effect of the gene that might be summarized in terms of a genetic relative risk (the ratio of age-specific incidence rates) or the absolute penetrance functions themselves, including the mode of inheritance (dominant, additive, recessive, or codominant). Next, one might be interested in the joint effects of the gene with other genes ($G \times G$ interactions), with environmental risk factors ($G \times E$ interactions), or the modifying effect of age or other host factors; such interactive effects are arguably most naturally tested in terms of relative risks, but ultimately it is the entire age-specific penetrance function with all its modifiers that we seek to estimate. Finally, we might be interested in the population attributable risk and in the attributable risk in various subgroups (e.g., by family history or ethnic group), quantities that involve both gene frequency and penetrance.

Each of these objectives must be studied in a particular context, depending on what is already known about the gene in question. This context will determine the feasibility and appropriateness of various design options. Some aspects of context include the following:

1) The nature of the outcome (quantitative, dichotomous, or age at onset); in studies of cancer, the end point is most appropriately viewed as censored age-at-onset data.
2) Mode of inheritance.
3) Whether the gene has a single mutant allele or is highly polymorphic.
4) Whether the mutant allele(s) are rare or common.

Correspondence to: Duncan C. Thomas, Ph.D., Department of Preventive Medicine, University of Southern California, 1540 Alcazar St., CHP-220, Los Angeles, CA 90033-9987 (e-mail: dthomas@rcf.usc.edu).

© Oxford University Press

5) Whether the main effect of the gene is strong or weak.
6) Whether there are other known genes or environmental factors with strong effects.
7) Whether there is a single genetic hypothesis under study or a large number of loci or even a genome-wide association scan.

Available Study Designs

We begin with the basic premise that estimation of population parameters (any of those discussed above, such as penetrance or allele frequency or their modifiers) requires some form of population-based study. Of course, depending on the sampling design, this may lead to study samples that are not representative of the entire population (e.g., samples of families with multiple cases), but the samples must ultimately be referable to a population base by means of a well-characterized sampling process if unbiased estimates of population parameters and valid tests of statistical hypotheses are to be possible. With this proviso, a wide range of study designs are available for addressing these aims, but most can be viewed as variants of the usual epidemiologic cohort and case–control designs for studying disease risks in relation to measurable risk factors. Zhao et al. *(1,2)* provide an overview of alternative study designs and their corresponding likelihoods.

For studying rare diseases (like cancer), there are compelling advantages to some form of case–control design, since a general population cohort study would require a very large cohort or long periods of follow-up, or both, to accrue sufficient cases. However, many of the genes of interest (like BRCA1) are also rare, so that an unselected series of cases may not yield an adequate number of mutation carriers. These two considerations suggest that neither the standard case–control nor the cohort designs will be optimal, and some form of multistage sampling will be called for. Furthermore, some of the more interesting designs will involve families, requiring consideration of the subtleties of dependent data, unlike the usual epidemiologic studies of independent individuals.

Case–Control Designs

The basic case–control design involves a comparison between case patients with the disease under study and control subjects who are free of the disease (at least at the time the case occurred) in terms of possibly causal factors (here the gene under study and its possible modifiers [host factors, other genes, and environmental factors]). Case patients and control subjects are frequently matched on confounding factors (such as age, sex, and race) that are not of primary interest. The two basic variants of this design are distinguished by their choice of control subjects: randomly selected from the source population of the case patient (the general population or subpopulations thereof) or family members of the case patient. Among family members, siblings or cousins are natural choices because control subjects should be roughly matched on age if environmental factors are also under consideration. An alternative involves the use of parents, not as control subjects *per se* but as a means of generating the distribution of possible genotypes the case patient could have inherited from the parents (what we shall call below "pseudosibs"). The relative merits of these alternatives are discussed by Witte et al. *(3)* and in the papers by Caporaso et al. *(4)* and Gauderman et al. *(5)* in this volume.

Cohort Designs

Whereas the case–control methodology is inherently limited to study of a particular disease, the cohort methodology allows direct observation of *all* of the end points that may be associated with a particular genetic syndrome, a particularly important consideration for genes whose effects may be pleiotropic. The need for long-term follow-up for chronic diseases like cancer can sometimes be alleviated by using a retrospective cohort study design, but for genetic studies, this study design would require availability of stored biologic samples for genotyping. As discussed by Langholz et al. *(6)* in this volume, more than one million individuals with stored samples have been enrolled in several large cohort studies dating back one or two generations that provide a rich resource for gene characterization studies. Most cohort studies of genetic factors to date have, however, used a familial cohort design, as discussed by Gail et al. *(7)* in this volume.

In a case–control study of familial aggregation, one might obtain family history information on case patients and control subjects and treat this information as a risk factor on which to compare case patients and unrelated population control subjects. Conversely, it is also possible to view the family members as a cohort at risk, classified in terms of their "exposure" to the disease status of the sampled case patient or control subject (whom we shall hereafter call the "proband," irrespective of whether affected or not). In a gene characterization study, the comparison of family members' outcomes is made in terms of measured genotypes, either their own or the proband's. The primary distinctions between the design alternatives are whether the family members' genotypes are also available and how the probands have been sampled.

Perhaps the simplest design is a study of disease incidence in first-degree relatives of genotyped probands as a cohort study. Wacholder et al. *(8)* use the term "kin-cohort design," and Gail et al. *(9)* use the term "genotyped proband design." Here, the proband might be either affected or unaffected, but if the mutation is rare, affected probands will have a higher yield of mutation carriers. [If the set of probands comprises both cases and unrelated population controls, this design is sometimes called a "case–control family study" *(1,10)*, although it is analyzed as a cohort study.] In this design, the relatives themselves are not genotyped, but the incidence is compared between relatives of carriers and noncarriers. Because roughly 50% of first-degree relatives of carriers will also be carriers and relatives of noncarriers will have roughly the population carrier frequency, it is possible to estimate the age-specific penetrance functions for a dominant gene from such a comparison *(9,11,12)*. Analogous calculations for a recessive gene could also be devised.

This basic design could be extended to include more distant relatives, using maximum likelihood methods to fit a form of segregation analysis model involving penetrance and allele frequency parameters, conditional on the proband's genotype and mode of ascertainment *(9)*. Allowance for the residual dependencies between family members because of sharing of other genetic or environmental risk factors is necessary to avoid bias *(7)* and obtain valid confidence intervals. A more direct estimate of penetrance is possible if family members' genotypes can also be obtained *(7,9)*. The cohort members can then be compared in terms of their own genotypes rather than their probands' genotypes and penetrance estimated using elementary methods of cohort analysis (except for the allowance for residual familiality

to obtain valid confidence intervals). More research is needed about the optimal selection of additional family members to genotype in these designs—whether it is better to genotype affected or unaffected family members, parents or siblings, near or distant relatives, for example; these issues are discussed in this volume in the context of the National Cancer Institute's Cooperative Family Registry for Colorectal Cancer Research *(13,14)*.

If a familial cohort study is to involve collection of extended families, then it is essential that one either select a fixed pedigree structure in advance (e.g., ascertain all family members of probands out to second degree) or follow the rules of sequential sampling of pedigrees *(15)* to avoid bias. Specifically, at any stage of the pedigree-building process, it is legitimate to use the data that have already been collected to decide whether to continue extending the pedigree, and it is also legitimate to use knowledge of the structure of the family at risk in a proposed branch of the family. However, use of anecdotal information about the existence of additional cases in that branch would lead to upwardly biased estimates of penetrance and allele frequencies. Furthermore, all of the data that are collected in this systematic fashion must then be included in the analysis (irrespective of whether additional cases were found). These rules were established for the purpose of segregation analysis, but they are equally applicable to familial cohort studies aimed at characterizing measured genes.

Multistage Sampling

Whether using a case–control or a cohort approach, the statistical efficiency of a study may be substantially improved by sampling families in two or more stages, based on a surrogate for genotype like family history *(13,16)*. For example, Whittemore and Halpern *(16)* describe a three-stage sampling design for a case–control family study of prostate cancer. In the first stage, case patients and population control subjects are selected and asked about the prevalence of prostate cancer in their fathers and brothers; in the second stage, a sample of case patients and control subjects with and without a family history (FH) are subsampled and extended-family histories are obtained and medically confirmed; families with three or more cases are then entered into the third-stage sample from which blood samples are drawn for genotyping. The question then arises as to how to select the sampling fractions at the second stage to minimize the variance of the parameter estimates (penetrances or allele frequency) or to maximize the yield of carrier families for a linkage analysis. For this design and a particular choice of parameters, they show that the optimal choice for parameter estimation is to sample 100% of case FH+ families, 34%–64% of control FH+ families, 12%–20% of case FH– families, and 3%–13% of control FH– families. A similar approach can be taken to optimizing the design of case–control studies with population control subjects *(17)* or family control subjects *(5)* or familial cohort studies with only the proband genotyped. For example, the efficiency of familial case–control designs for estimating the main effect of a genotype or a $G \times E$ interaction effect can be substantially improved by restriction to multiple-case families in various ways.

A limitation of multistage sampling is that it tends to focus on only a single disease outcome, whereas many cancer-predisposing conditions cause multiple cancers to occur in families. A sampling design that is optimized to select families that would be most informative for a given cancer may then be inefficient for studying these other end points associated with the same genes. This issue also arises in the analysis as a need to account for competing risks, as in the analysis of breast and ovarian cancers by Whittemore et al. *(12)*. Finally, any comparisons of relative efficiency must carefully account for the different costs at each of the stages of sampling.

Use of High-Risk Families

Pedigrees with many cases are highly informative for linkage analysis but typically have not been ascertained in any population-based manner. Frequently, such families have initially come to the attention of a genetics clinic by virtue of having an unusual number of cases and are then perhaps further extended in a fashion that may or may not respect the rules of sequential sampling of pedigrees *(15)*. Even if such further extension has been done systematically, the initial cluster may be difficult to define retrospectively as is necessary for ascertainment correction. Obviously, estimates of penetrance and allele frequency from such families will tend to be seriously overestimated unless the ascertainment and extension processes are taken into account, and, in most such instances, it is virtually impossible to characterize these processes in terms of any statistical sampling design. Nevertheless, it is theoretically possible to analyze the data in terms of the "retrospective" likelihood of the genotypes given the observed phenotypes [*(18,19)*; Kraft P, Thomas DC: submitted for publication] to produce consistent estimates of population parameters; for genes with high penetrance, such estimates based on linked markers (the "MOD score" approach) are generally thought to be fairly efficient but not for genes with low penetrance. A more fundamental problem, however, is that this approach implicitly assumes that the penetrance (or genetic relative risk) is homogeneous across families. If it is not, say, because of genetic heterogeneity due to other genes or environmental factors, then the average penetrance estimated from high-risk families will tend to be different from that estimated from a population-based series of families [*(7,9)*; Kraft P, Thomas DC: submitted for publication]. For example, high-risk families may be segregating other (unknown) genes, some of whose effects could be incorrectly ascribed to the gene under study if this possibility is not explicitly addressed in the model. Further research is needed to study the potential biases and relative efficiency of designs based on high-risk families with the use of the retrospective likelihood approach.

Case-Only Designs for Gene–Environment Interactions

All of the above designs can be used for testing gene–environment interactions, as well as the main effect of a gene, as discussed by Goldstein and Andrieu *(20)*. However, one additional design is available for this purpose that is not amenable to studying the main effeets of genetic or environmental factors—the so-called "case-only" or "case–case" design *(21–23)*. Under an assumption that genotype and an environment factor are independently distributed in the source population, one can interpret an association between genotype and exposure among cases as evidence of a gene–environment interaction. This design thus entails ascertainment of cases only and comparing the distribution of environmental factors between carriers and noncarriers.

Statistical Analysis Issues

Each of the above designs raises a host of statistical analysis issues that are generally beyond the scope of this paper. A few

key observations are worth pointing out, however. First, the analysis must respect the statistical sampling design, e.g., a matched case–control design would normally require a matched form of analysis (e.g., conditional logistic regression); a multistage sampling design must allow for the sampling fractions at each stage. Second, the form of analysis depends on the parameters of interest: A standard analysis of case–control data (whether using population or family controls) allows a direct estimate of the genetic relative risk and its modifiers, but estimation of absolute penetrances would require either external information (such as population rates) or some form of cohort design. Third, the bias and statistical efficiency of a design can only be defined in relation to a particular form of analysis, since some designs may allow more than one form of valid analysis. Finally, maximum likelihood analysis of a model that is correctly specified (both in terms of the true state of nature and the study design) will generally be asymptotically unbiased and fully efficient but may be more sensitive than other kinds of analysis to model misspecification. For example, it is likely that an analysis that ignores an important interaction or another important gene will generally lead to biased results whatever the design. Conversely, certain types of misspecification, such as residual familial aggregation, might be overcome by using robust "estimating equations" methods *(24)* or regressive models *(25)*.

Discussion

Our aim in the subsequent papers in this panel *(4–7,13,20)* is to provide some guidance about the best choice of design for addressing the various aims in the various contexts described at the outset. In the following section, we set up a general framework for addressing this problem and attempt to summarize what we believe is already known. To orient the reader to this discussion, we now lay out what we see as the "Big Questions."

1) How can one determine whether a particular design will lead to unbiased estimates of a particular population parameter? For example, under what conditions might a case–control design with unrelated population controls lead to bias? In what ways might family-based, case–control or familial cohort studies be biased?
2) Are there important differences in power or statistical efficiency between designs? Is it possible to optimize the sampling plan for a particular objective?
3) Given a selection of a basic design, how should one decide whom to genotype?
4) What practical or other nonstatistical issues need to be taken into account?
5) What use, if any, can be made of families not selected in a population-based manner (e.g., heavily loaded pedigrees selected for linkage analysis) for estimating population parameters such as penetrance?
6) Once a gene has been identified, what are the best approaches to characterize other genes or environmental factors, or both?
7) Do the types of study depend on the types of gene involved, e.g., high versus low penetrance, common versus rare alleles?
8) Is it feasible to develop a population-based design that would be efficient both for characterization and for gene mapping studies?

Bias

Perhaps the question an investigator must consider first is whether to use unrelated or familial controls in any of the designs discussed above. To focus the discussion of this question, let us consider the case–control designs. Assuming the starting point is a population-based series of case patients, the primary concern about bias centers on the source of control subjects; both population control subjects and family control subjects are potentially subject to bias, but in quite different ways.

Population control subjects are particularly prone to a form of confounding bias known as "population stratification" *(4)*, a situation that arises when both the gene frequency at the locus under study and the penetrances vary between subpopulations that are not accounted for by matching, stratification, or adjustment. Such confounding can arise through the effects of other genes, host factors, or environmental exposures. Suppose that the gene under study in fact has no causal effect but is associated in the population with some other gene that does have a causal effect; if the other gene is unknown or unmeasured, a spurious association with the gene under study will be seen. Such gene–gene associations can arise even in homogeneous populations through linkage disequilibrium with *linked* genes. Such associations are of considerable interest for the purpose of gene mapping but, from the perspective of gene characterization, are viewed as spurious. Gene–gene associations can also arise between *unlinked* genes in ethnically diverse populations simply as a result of a mixture of different allele frequencies at the two loci and are of no interest whatsoever. Likewise, it is possible for confounding by nongenetic risk factors to occur in a mixture of subpopulations with different prevalences of the two risk factors. As in any other form of epidemiologic study, confounding is controlled by matching or by stratified analysis, but these approaches require that the confounder be identifiable. Control of confounding can be very difficult for unknown genes that could have powerful risks and strong gradients between or even within ethnic groups, especially in freely mating multiethnic populations. For example, it may be difficult to determine the relevant ethnic subgroup for a Caucasian subject (e.g., northern or southern European), and finding a suitable control subject for a case patient with grandparents from four different ethnic groups may be virtually impossible.

Several examples of spurious gene associations thought to be due to population stratification have been documented, such as the D_2 allele of the dopamine receptor gene with alcoholism *(26,27)* and the $Gm^{3;5,13,14}$ immunoglobulin haplotype with non-insulin-dependent diabetes among Pima-Papago Indians *(28)*. Although these associations may be isolated examples, the magnitude of genetic risks and the frequent observation of large gradients between and within ethnic groups suggests that such confounding could be a much more serious problem than in other types of epidemiologic studies.

The problem of population stratification is completely eliminated by the use of siblings or pseudosibs as control subjects because they are descended from identical gene pools. Cousins do not provide such absolute control because of the possibility of intermarriage but, to the extent that people tend to marry within their ethnic groups, will generally avoid population stratification bias as well. Drawing cousin controls from both branches of mixed-race families can also help, but it can also introduce subtler biases *(3)*.

Family control subjects are not immune to other forms of

bias, however. First, the pool of potential relative control subjects is limited and not every case may have a suitable control. If availability of a control is itself a risk factor (directly or indirectly, say by socioeconomic correlates of family size), then selection bias could result. Second, matching on age will generally be more difficult, again because of the limited number of relatives available. Control subjects should be required to have attained the case patient's age at diagnosis and still be disease free, but this situation will generally lead to drawing control subjects from older relatives, leading to possible confounding by secular trends. Third, if multistage sampling is used to focus on multiple case families, then any restriction on case patients must apply equally to control subjects to avoid bias. This restriction is generally easy for sibling or pseudosib control subjects but not for cousins: If the case patient is required to have an affected mother or sibling, a sib control subject meets this requirement automatically, but a cousin would also have to have an affected mother or sibling with the same relationship to the case patient (i.e., if the case patient's mother is the aunt of the cousin, the cousin's mother would have to be the aunt of the case patient) *(5)*.

Finally, we have alluded earlier to bias from model misspecification, in particular to analyses of familial cohort designs that ignore residual familial aggregation not accounted for by the measured genes or by other risk factors, or to phenotypic dependencies between family members. The standard conditional analysis of case–control studies avoids this difficulty by eliminating all stratum parameters (e.g., family- or ethnic-group specific baseline risks or allele frequency parameters), leading to unbiased estimates of the genetic relative risk. This option is not available for some types of familial cohort designs, and simply adding additional nuisance parameters to the model for each family is not a viable solution. A random-effects model may offer a valid analysis, but only if the form of that model is correctly specified *(Kraft P, Thomas DC: submitted for publication)*.

Statistical Efficiency

By statistical efficiency, we refer to the precision of the parameter estimates from a particular study design and analysis combination. Generally, efficiency is described in terms of asymptotic relative efficiency (ARE) comparing the large sample variances of the estimated parameters from two alternative designs. The ARE can be interpreted as the ratio of sample sizes required to attain the same degree of statistical precision, and it is also closely related to the power of the test to reject a null hypothesis against some alternative. The large sample variances are computed from the inverse of the expected Fisher information (the negative of the inverse matrix of second derivatives of the log likelihood), averaged over all possible study outcomes (diseases and genotypes in a family) conditional on the study's sampling plan and a particular set of model parameters. One must appreciate, however, that the ARE captures *only* the statistical sampling variability, not any of the other qualitative uncertainties or the relative costs of design alternatives.

As an example of relative efficiency calculations, consider the choice between use of population, sibling, pseudosib, or cousin control subjects in a case–control design for estimating relative risks. Without population heterogeneity (or matching on ethnicity), genotypes of unrelated controls would be independently distributed from those of the cases, whereas familial controls would tend to be associated with them, more closely for siblings than for cousins. This association would suggest that sibling control subjects would tend to be somewhat "overmatched" compared with population control subjects, leading to a less efficient design, whereas cousins would have intermediate efficiency. This intuition is correct, but the magnitude of the AREs depends on the true genetic model: For a multiplicative or a dominant gene, the sib control design has roughly 50% efficiency and the cousin control design 88% efficiency, relative to population control subjects; for a recessive gene with relative risk (RR) = 20 and allele frequency of 14%, these figures rise to 69% and 97%, respectively. The pseudosib design, however, is always at least as efficient as population controls (101%–106% for dominant and multiplicative models) and much more efficient (ARE = 231%) for major recessive genes *(3,5)*.

Siblings (and to a lesser extent cousins) would also tend to have similar environmental exposures. Thus, one might expect that they would tend to be even less efficient for testing $G \times E$ interactions, but this intuition is not always correct *(3)*. To test a twofold interaction RR, sib control subjects can be *more* efficient than any of the other four designs (ARE = 214%, 136%, and 275% for major recessive, multiplicative, and dominant genes, respectively), depending on such additional factors as the exposure prevalence (here 50%), the exposure concordance within families (odds ratio = 2), and the main effect of the environmental factor (RR = 2). The reason for this greater efficiency of sibling controls is that the most informative pairs are those that are genotype concordant and exposure discordant, and the proportion of such pairs is higher for sibling controls by virtue of their genetic relationship. Genotype-discordant, exposure-concordant, and jointly discordant pairs are also informative but are less common *(3,5)*.

Such calculations are easily performed for a variety of design and analysis combinations and form the basis for power calculations for designing studies. Examples are discussed elsewhere in this volume for family-based, case–control *(5)*, kin-cohort *(7)*, and multistage *(13)* designs. To date, no general purpose software for design calculations presently exists, but this is an aim of the Informatics Center for the Cooperative Family Registries for Breast and Colorectal Cancer Research (Dr. Hoda Anton-Culver, principal investigator).

Practical Issues

The choice of a preferred design must, of course, turn on more than purely statistical considerations. We have already alluded to some of the potential difficulties in obtaining appropriate controls for a case–control study with either population or family controls—the difficulty of ethnic matching for the former, problems of availability and age-matching for familial controls, and family-history matching for multistage designs. Cooperation is another dimension: It has often been the experience of epidemiologists that family members are more highly motivated to participate than are randomly selected members of the population, but more difficult ethical issues arise when studying families. Families are frequently geographically dispersed, posing substantial logistical difficulties and additional costs. Comparisons of cost-efficiency must take account of the different types of costs—per family, per subject, per genotype, etc. The choice of design will also be influenced by such factors as survivorship: The pseudosib design is unlikely to be viable for such cancers as prostate because most parents will already be

dead, and the familial cohort designs with genotyping of family members may run into difficulties if most affected relatives are dead. The latter design also requires a carefully thought-out protocol for the selection of family members to genotype to avoid bias because of nonrandom selection and to maximize the information yield.

How Serious Is the Population Stratification Concern?

The potential bias because of confounding by ethnicity was discussed above and in greater detail by Caporaso et al. *(4)*. As noted there, several textbook examples of spurious associations that appear to be because of population stratification have been documented, but the extent of the problem remains unknown. One other example of a different nature is worth mentioning. Insulin-dependent diabetes mellitus has consistently been found to be associated with a polymorphism (5′FP) near the 5′ end of the insulin gene *(29)*, but affected sib pair studies have found no evidence of linkage *(30)*. However, the transmission-disequilibrium test (TDT), a variant of the pseudosib design, was able to reject the joint null hypothesis of no association or no linkage *(31)*. This situation appears to be an instance of a true population association that might have been incorrectly attributed to population stratification and could only be established by an appropriate family-control design. Here, it appears that the pure linkage tests failed because of their low power for detecting genes with only moderate effects.

How can we move beyond such anecdotal examples to a more systematic investigation of the population stratification problem? The difficulty is that truly causal genes for most diseases are probably still unknown, so we cannot directly observe the gene–gene associations that would resolve the issue. However, we do have ethnicity as a surrogate. Many diseases are known to have strong gradients (sometimes more than 100-fold) in incidence between ethnic groups, and many genes likewise show strong gradients in allele frequencies across ethnic groups. What is not known is how much of the gradient in incidence of a particular disease is directly attributable to gradients in frequency of already known genes and how much may be because of other factors (genetic or environmental) with which they may be associated. "Ecologic" studies with careful characterization of ethnic origin (as finely as possible, at least back to grandparents) might shed useful light on this question. Parallel family-based and population-based, case–control studies might also be useful, but they would require careful ethnic characterization to interpret any differences in conclusions. Studies might also be designed to take advantage of knowledge of allele frequencies in ancestral populations or to use a panel of polymorphic markers to infer ethnic origin *(32,33)*.

What Are the Merits of Integrated Designs?

Zhao et al. *(1,2)* have suggested that it may be feasible to marry the objectives of gene discovery and gene characterization by a global research program firmly anchored in the principles of population-based, multistage sampling design for family studies. We have seen in this panel how these principles can indeed be put to good use to design unbiased and highly efficient studies for characterizing already identified genes. It is also evident that multistage sampling will allow heavily loaded pedigrees to be identified and extended in a manner that still allows estimation of population parameters. There are two outstanding questions, however. First, is this an *efficient* way to identify families that are highly informative for linkage analysis, compared with more anecdotal sources, such as high-risk cancer family clinics, and can any use be put to high-risk families ascertained in unsystematic ways for the purpose of estimating population parameters? Second, is it possible to develop optimal study designs for resources that are aimed at both purposes. For example, Whittemore and Halpern *(16)* note that the optimal designs for parameter estimation require all strata (defined by case–control status of the proband and family history) to be sampled with nonzero probabilities, whereas the optimal designs for linkage analysis (based on the yield of carrier families) entails enrolling *all* families from the highest prevalence strata until the desired sample size has been attained. These two objectives are thus fundamentally incompatible, and no design would be optimal for both purposes. The choice of a reasonable compromise would thus turn on the subjective importance of the two aims to the investigator. Further research along these lines would be helpful. Whittemore *(34)*, Zhao et al. *(2)*, and Haile et al. *(14)* in the Integration Panel of this volume address these questions further.

How Do We Decide Which Design to Use?

Several papers have addressed various aspects of these design trade-offs. In the author's view, there are compelling advantages to the use of some form of family design for characterizing measured genes over case–control or cohort studies of independent individuals, most importantly because of their avoidance of the potential bias from population stratification. However, we recognize several difficulties with this approach, both statistical and practical. Within the class of family studies, there are a large number of choices, both in the selection of an overall study design as well as in the selection of specific family members to genotype. Although much is already known about the optimal choice of a sampling plan within a limited range of designs (e.g., choice of controls in case–control studies; choice of sampling fractions in multistage family cohort studies), there has been no systematic comparison of the full range of designs across the full spectrum of aims and contexts described at the outset. If we are ultimately to be able to address the question of what design is the best for a particular study, a much more systematic investigation of these design issues will be needed.

References

(1) Zhao LP, Hsu L, Davidov O, Potter J, Elston R, Prentice RL. Population-based family study designs: an interdisciplinary research framework for genetic epidemiology. Genet Epidemiol 1997;14:365–88.

(2) Zhao LP, Araguki C, Hsu L, Potter J, Elston R, Malone KE, et al. Integrated designs for gene discovery and characterization. Monogr Natl Cancer Inst 1999;26:71–80.

(3) Witte J, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case–control studies of candidate genes and gene–environment interactions: basic family designs. Am J Epidemiol 1999;149:693–705.

(4) Caporaso N, Rothman N, Wacholder S. Case–control studies of common alleles and environmental factors. Monogr Natl Cancer Inst 1999;26:25–30.

(5) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Monogr Natl Cancer Inst 1999;26:31–7.

(6) Langholz B, Rothman N, Wacholder S, Thomas DC. Cohort studies for characterizing measured genes. Monogr Natl Cancer Inst 1999;26:39–42.

(7) Gail MH, Pee D, Carroll R. Kin-cohort designs for gene characterization. Monogr Natl Cancer Inst 1999;26:55–60.

(8) Wacholder S, Hartge P, Struewing JP, Pee D, McAdams M, Brody L, et al. The kin-cohort study for estimating penetrance. Am J Epidemiol 1998;148: 623–30.

(9) Gail MH, Pee D, Benichou J, Carroll R. Designing studies to estimate the

penetrance of an identified autosomal dominant mutation: cohort, case–control, and genotyped-proband designs. Genet Epidemiol 1999;16:15–39.
(10) Whittemore AS. Logistic regression for family data. Biometrika 1995;82: 57–67.
(11) Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, et al. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 1997;336:1401–8.
(12) Whittemore AS, Gong G, Intyre J. Prevalence and contribution of BRA1 mutations in breast cancer and ovarian cancer. Results from three U.S. population-based case–control studies of ovarian cancer. Am J Hum Genet 1997;60:496–504.
(13) Siegmund KD, Whittemore AS, Thomas DC. Multistage sampling for disease family registries. Monogr Natl Cancer Inst 1999;26:43–8.
(14) Haile RW, Siegmund KD, Gauderman WJ, Thomas DC. Study-design issues in the development of the University of Southern California Consortium's Colorectal Cancer Family Registry. Monogr Natl Cancer Inst 1999;26:89–93.
(15) Cannings C, Thompson EA. Ascertainment in the sequential sampling of pedigrees. Clin Genet 1977;12:208–12.
(16) Whittemore AS, Halpern J. Multi-stage sampling in genetic epidemiology. Stat Med 1997;16:153–67.
(17) Langholz B, Goldstein L. Risk set sampling in epidemiologic cohort studies. Stat Sci 1996;11:35–53.
(18) Risch N. Segregation analysis incorporating linkage markers. I. Single-locus models with an application to type I diabetes. Am J Hum Genet 1984;36:363–86.
(19) Hodge S, Elston R. Lods, wrods, and mods: the interpretation of lod scores calculated under different models. Genet Epidemiol 1994;11:329–42.
(20) Goldstein AM, Andrieu N. Detection of interaction involving identified genes: available study designs. Monogr Natl Cancer Inst 1999;26:49–54.
(21) Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case–control studies. Stat Med 1994;13:153–62.
(22) Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene–environment interaction: case–control studies with no controls! Am J Epidemiol 1996;144:207–13.
(23) Umbach DM, Weinberg CR. Designing and analyzing case–control studies to exploit independence of genotype and exposure. Stat Med 1997;16: 1731–43.
(24) Zhao LP, Hsu L, Holte S, Chen Y, Quiaoit F, Prentice RL. Combined association and aggregation analysis of data from case–control family studies. Biometrika 1998;85:299–315.
(25) Bonney G. Regressive models for familial and other binary traits. Biometrics 1986;42:611–25.
(26) Gelernter J, Goldman D, Risch N. The A1 allele at the D2 dopamine receptor gene and alcoholism: a reappraisal. JAMA 1993;269:1673–7.
(27) Blum K, Noble EP, Sheridan PJ, Montgomery A, Ritchie T, Jagadesswaran P, et al. Allelic association of human dopamine D2 receptor gene in alcoholism. JAMA 1990;263:2055–60.
(28) Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. $Gm^{3;5,13,14}$ and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 1988;43:520–26.
(29) Bell GI, Horita S, Karam JH. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 1984; 33:176–83.
(30) Spielman RS, Baur MP, Clerget-Darpoux F. Genetic analysis of IDDM: summary of GAW5-IDDM results. Genet Epidemiol 1989;6:43–58.
(31) Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506–16.
(32) Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, et al. Ethnic-affiliation estimation by use of population-specific DNA. Am J Hum Genet 1997;60:957–64.
(33) Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 1999;65: 220–8.
(34) Whittemore AS, Nelson LM. Study design in genetic epidemiology: theoretical and practical considerations. Monogr Natl Cancer Inst 1999;26: 61–9.

Note

Supported in part by Public Health Service grants R01CA52862 (National Cancer Institute) and P30ES07048 (National Institute of Environmental Health Sciences), National Institutes of Health, Department of Health and Human Services.

Case–Control Studies of Common Alleles and Environmental Factors

Neil Caporaso, Nathaniel Rothman, Sholom Wacholder

It is clear from descriptive and migration studies that most cancer is environmental in origin. Descriptive, case–control and cohort studies have provided the foundation for our understanding of the environmental component of cancer etiology as well as most major causes of morbidity and mortality. We propose that the same epidemiologic methods that have provided fundamental insight into the etiology of cancer in the general population are optimally suited to study the impact of relatively common polymorphisms on chronic disease incidence. In this article, we describe the role of case–control studies in assessing the effects of genes in disease. Some of the advantages and disadvantages of the case–control design, particularly as an alternative to case–control studies nested in a cohort in the context of the study of complex disease, are described. [Monogr Natl Cancer Inst 1999;26:25–30]

Population-based epidemiologic studies including descriptive, case–control, and cohort studies have provided the foundation for our understanding of cancer etiology. The accumulated knowledge provided by epidemiologic research carried out during the last 50 years has unequivocally shown that the vast majority of cancer derives from environmental exposures, broadly defined. Descriptive epidemiology reveals striking differences in international cancer incidence rates. Migration studies demonstrate that rates can shift within one generation to the rates of the new country *(1)*. Case–control and cohort epidemiology studies have provided convincing evidence for causal relationships between cancer and smoking, alcohol, viral, drug, radiation, occupational, and lifestyle exposures *(2)*.

There is broad agreement that the relationship between environmental exposures and cancer risk is likely to be modified by genetic factors, since *(a)* most xenobiotics that enter the body and particularly human carcinogens undergo metabolic processing and *(b)* wide differences in these capacities between people are often due to heredity *(3,4)*. Many of the specific genes and respective substrates involved are known (Table 1), but the magnitude of the risks and understanding whether gene–environment or gene–gene effects are involved remain controversial. We propose that the same epidemiologic methods that have provided fundamental insight into the etiology of cancer in the general population are optimally suited to evaluate the impact of relatively common polymorphisms on chronic disease incidence as well as the potential interaction between these alleles and the environmental exposures that drive disease risk.

This article highlights the advantages of population-based studies. We do not question the central importance that linkage studies have played in elucidating the genetic origins of family cancer syndromes *(5)*. We note, however, that, for the vast majority of human cancer that exhibits familial tendency but not a simple mendelian inheritance pattern, the tools of classic cancer epidemiology may be better adapted to unraveling the joint effects of environment and genes. This article highlights the advantages of population-based studies in the investigation of common genetic variation in cancer and, in particular, focuses on case–control studies.

Cohort studies, considered in a companion article *(6)*, and case–control studies are sometimes referred to as "association" studies, implicitly reminding others of the truism that "association does not imply causation." Linkage studies, of course, also use statistical evidence to infer causation. Results from any study design need to be evaluated within the context of the historical *(7)* and ongoing *(8)* debate regarding the types of evidence that can advance an association to the level of causation.

Studies of genetic markers in population-based studies were responsible for identifying the relation of HLA-B27 and autoimmune disease, apolipoprotein E in late-onset Alzheimer's disease *(9,10)*, and ABO blood group and gastric cancer *(11)*. In the 1970s and 1980s, this approach was used to study putative associations of cancer with metabolic phenotypes *(12–15)*. By the early 1990s, technology allowed germline DNA to be used in study-specific genetic polymorphisms in relation to particular cancers *(16,17)* as indicated in Table 1. While the evidence for most of the associations is mixed or sparsely studied, the biologic and epidemiologic data generally support associations between NAT2 (the "slow" acetylation phenotype) *(18)* and bladder cancer and are at least suggestive for GSTM1 and some smoking related-cancers *(19–21)*. Plausible hypotheses for gene–human cancer/disease associations are the focus of active investigation based on the appreciation that genetic traits can influence the disposition of relevant exposures or mechanistically relevant features on the disease pathway (Table 1).

Case–Control Studies

The case–control design became widely recognized as an efficient and scientifically sound approach with the studies in the early 1950s linking tobacco smoking and lung cancer *(37)*. Since then, the cancer epidemiology community has accumulated extensive experience with population-based case–control studies. Because of the efficiency and the telescoped time required to complete a case–control study relative to a full cohort study, case–control studies have provided initial clues and been a major, if not the major, source of evidence for many established exposure–disease relationships. They can provide evidence for estimates of the relative and attributable risk and sometimes the absolute risk. In fact, the key weakness of case–control studies of environmental factors, reliance on self-report, does not arise

Affiliation of authors: Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

Correspondence to: Neil Caporaso, M.D., National Institutes of Health, Executive Plaza South, Rm. 7116, Bethesda, MD 20892 (e-mail: caporaso@nih.gov).

See "Note" following "References."

Table 1. Cancer, metabolic polymorphisms, and proposed mechanisms

Cancer/disease	Gene (reference No.)	Proposed basis
Lung	CYP1A1 *(22)* CYP2E1 *(23)* CYP2D6 *(24)*	Phase 1 genes that activate known tobacco carcinogens, including polycyclic aromatic hydrocarbons, *N*-nitrosamines, and aromatic amines
	GSTM1 NAT2 EPHX	Phase 2 gene involved in detoxification and elimination of carcinogenic epoxides or aromatic amines derived from tobacco
Nasopharyngeal	GSTM1 *(25)*	Detoxification of benzo[a]pyrene or other carcinogens in tobacco smoke
	CYP2E1 *(26)*	Activation of nitrosamines may influence risk
Oral	ADH *(27)*	Enhanced production of carcinogenic byproducts of alcohol metabolism
Gastric	CYP2E1	Activation of nitrosamine carcinogens
Hepatocellular	EPHX *(28)* GSTM1	Activation and elimination of aflatoxin B1
Esophageal	CYP2E1 *(29)*	Activation of nitrosamine carcinogens
Bladder	NAT2 *(30)*	Decreased elimination of aromatic amines in "slow acetylators"
Breast	CYP1B1	Metabolism of estradiol to catechol estrogens
	CYP17	Metabolism of steroids
	NAT2 *(31)*	Decreased detoxification of aromatic amines in smokers
Colorectal *(32)*	NAT2	Activation of food-borne carcinogens
Prostate	Androgen receptor *(33)*	Alter receptor transactivation with impact on androgen effect
	SRD5A2 *(34)*	Metabolic activation of testosterone to dihydrotestosterone
Non-Hodgkin's lymphoma	CCR5 *(35)*	In patients with human immunodeficiency virus, chemokine receptor defect alters risk of infection and development of acquired immunodeficiency disease syndrome-related cancer
Renal	CYP1A1 *(36)*	Phase 1 gene that activates tobacco carcinogens

in studies of constitutional risk factors. Case–control studies conducted in the setting of an established cohort study are described in an accompanying article *(6)*.

Biomarkers of Genetic Susceptibility

Modern epidemiologic studies that evaluate genetic factors include DNA-based genotype determinations. When genotype could be determined only by laborious metabolic phenotype determinations involving probe drug administration or complex *in vitro* laboratory studies, investigators encountered methodologic difficulties, including selection bias (i.e., only relatively healthy subjects could undergo phenotyping) and limited power (i.e., difficult laboratory assays or clinical requirements imposed time, cost, and labor constraints on study size). The difficulties of accommodating phenotype assays in field studies generally limited the scope to the characterization of single genes. Today, the genotype information is directly assessed by analysis of germline DNA, obtained from blood or increasingly from other sources (mouthwash rinse, buccal swabs, or stored tissue). Still, the requirement for biospecimens, though less arduous than that imposed during the phenotyping era, continues to have important implications for study design *(38,39)*, measurement error, and sample collection and processing *(40)*.

Beyond issues that derive from the need to obtain biomarkers on study subjects *(41)*, the study of genetic factors imposes further considerations. One issue is that the exploding genetic information and technology will result in multiple and posteriori comparisons. A particular issue that is seemingly unique to associations with genetic markers is the problem of population stratification, considered briefly in the next section.

Population Stratification

Population stratification is a consequence of different rates of disease in different ethnic groups; any genetic or environmental factor whose distribution differs between ethnic groups may appear to be related to disease even though there is no causal relationship. In the simplest situation, if there is a difference in risk between two ethnic groups in the study base, the ethnic groups will be distributed differently between the case and control groups so that any genetic (or other environmental) factor that differs between ethnic groups will tend to be associated with the risk of disease. The worst case for population stratification occurs when two separate populations with widely different allele frequencies and disease rates are admixed as described by Knowler et al. *(42)* in the study of diabetes and the Gm haplotype in the Pima Indians. An empirical investigation of the extent of bias from the slow acetylation genotype or phenotype and male bladder or female breast cancer found very little bias in a study base of non-Hispanic European-Americans in a multiethnic setting in the United States (Wacholder S et al.: submitted for publication, 1999). We further showed that, when multiple populations are incorporated, as is likely in most U.S. case–control studies, confounding will be minimal. Controlling crudely for ethnicity will often (though not always) control some of the remaining bias. We, therefore, disagree with those who have stated that population stratification undermines the conclusions from population-based studies involving candidate genes so pervasively that familial controls are required to rule out this problem *(43,44)*. The key question is whether a population-based case–control study is so severely affected by this bias that the results are not credible or whether, as we believe, the bias is small and tolerable.

Even if population control subjects have general utility, eliminating residual concerns about population stratification, identifying gene–environment interactions *(45)*, and efficiency considerations may support the use of relative control subjects for a particular study. The relative efficiency of using relatives as control subjects against unmatched population control sub-

jects is considered in another article *(46)* in this volume. The cost efficiency of selecting an appropriate population control subject and a parent or relative will affect the statistical efficiency as a function of factors such as age, residence of relatives, and the availability of potential rosters. For older control subjects, parents and siblings may be deceased; already identified case subjects will need to be excluded because no eligible control subject is available. Relatives may live far away; extra cost to locate and visit them will be incurred, unless self-administered DNA collection, such as a buccal cell swab, and questionnaire are used. We suspect that, given the added complexity of choosing relative control subjects, the total cost will be higher than that for population control subjects. Unwarranted concern about population stratification, even when it is likely to be trivial, can render the costs of epidemiologic studies prohibitive.

The Case–Control Method in Relation to Other Epidemiologic Designs

Population-based case–control studies have had a crucial role in unraveling complex human diseases *(47–51)*. Now, however, existing cohort studies that have or are collecting blood samples or other sources of genomic DNA will accrue large numbers of case subjects with common tumors over the coming years. Appropriately, questions are being raised about the utility of carrying out new case–control studies, either population based or hospital based, to study the main effects of common polymorphisms and their interaction with environmental exposures. Designers of a new case–control study will need to show that it offers benefits that cannot be obtained from existing cohorts already under study. We, therefore, raise some considerations when planning to carry out a new case–control study involving genetic factors in contrast to performing nested studies within existing cohorts.

Considerations Before Launching a New Case–Control Study

Tumor Incidence

Perhaps the key advantage of case–control studies is the ability to enroll relatively large numbers of cases of the less common tumors. Given the need for sample sizes up to several thousand cases and controls to study gene–environment interaction *(52,53)*, it is only feasible to collect enough cases of the more common tumors in most cohort studies without pooling across cohorts.

Inclusion of Diverse Population Groups

Case–control studies can focus on enrolling a narrow range of ethnic and racial features, age, or socioeconomic levels that is particularly interesting or important but not adequately represented in existing cohort studies.

Exposure

Depth of exposure data. Case–control studies can collect more detailed and broader information about exposure from both interviews and records than is feasible in a cohort study. This is particularly important when there is concern about a specific type of exposure that is not generally assessed at all or in adequate detail in the typical cohort questionnaire (which usually focuses on diet and general lifestyle factors). Examples could include occupational and environmental exposures requiring complete occupational and residential histories, respectively. Cohorts have an inherent limitation, in that, having as their aim the study of multiple end points, they can collect less extensive data on exposures relevant to any one particular disease, although the opportunity to return to participants at later time points may partially ameliorate this point. Case–control studies can rapidly respond to and focus on new exposures that are currently of concern for particular tumors, tailoring methods to optimally capture target data. In contrast, cohort studies will have instruments in place that will inevitably lack precision or entirely miss new exposures.

Retrospective versus prospective exposure determination. Studies that rely on retrospective exposure assessment that may be affected by disease or its treatment or on questionnaire responses susceptible to rumination by respondents are liable to bias from differential misclassification. Biomarkers (except germline DNA) and responses to questionnaires may change as a consequence of the early disease process or diagnosis itself. We have previously shown a striking drop-off in power to detect gene–environment interactions occurring under some circumstances with even small to moderate exposure misclassification *(38,53)*. Furthermore, even with very rapid case ascertainment, case mortality is a problem for the study of very aggressive tumors in population-based case–control studies. The reliance on prospective exposure assessment preceding the onset of disease protects against differential misclassification. Biochemical measures of exposure are more suitable in cohort settings if the time of collection (i.e., not influenced by preclinical disease) and biologic characteristics (i.e., not influenced by random within-person variation) are appropriate. In addition, case–control studies are usually limited to one retrospective measurement of exposure, whereas cohort studies may conduct repeated measures over time. Although it is feasible to analyze genetic polymorphism in tumor blocks, we have found that the proportion of samples that can be successfully studied is highly variable across studies. Tissue availability may vary by factors related to genetics; i.e., availability of tissue may be limited to "early" surgical cases.

Gene–environment interactions. The strong focus on collection of exposure data is well suited to evaluate evidence for interaction between genes and exposure *(54)*. A number of "variant" case–control designs have been proposed to facilitate the study of interaction, particularly the case-only or case-case design *(55–57)*. Subject to certain assumptions (i.e., exposure and gene are independent) and limitations (i.e., the independent effects of gene and exposure cannot be assessed; generally but not always limited to departures from multiplicativity), this approach may be applicable to both case–control studies and nested studies.

Case and Control Selection

Ascertainment. A major concern of case–control studies is proper case and control selection. Proper control subjects are representative of the study base from which the case subjects arise *(58,59)*. Identifying either a random sample from the general population or the source population for case subjects presenting at a particular hospital(s) may be difficult.

Participation. Further potential for selection bias occurs if case or control subjects are less likely to participate because of

problems in the collection of biospecimens. Since the source population for cohorts is explicit, selection bias is less of a problem as long as follow-up rates are high *(60)*. Low participation rates in case–control studies and, particularly, refusals related to providing DNA can bias results, especially when case subjects are less likely to participate than control subjects and selection is related to the gene. Low participation rates also threaten the population-based nature of the study, undermining its use for estimating absolute and attributable risks *(60)*. A promising solution to low participation rates for phlebotomy is the less invasive collection of buccal cells *(61)*.

Collection of Biospecimens

Case–control studies, particularly when hospital-based, offer distinct advantages for the collection of biospecimens. It is not feasible for a large cohort study to reach newly incident cases at a widely dispersed set of hospitals. A case–control study in a narrow geographic region or with the use of only a few hospitals can establish an efficient collection of blood, urine, cells, surgical tissue, and pathologic material along with supporting documentation (medical records). Although rapid contact is not important for germline-derived genetic markers because they are invariant to the presence of disease in the host, case–control studies also have an advantage if fresh tumor tissue is a study requirement. Resources for collection, processing, and storage are often conveniently available around the hospital setting. Often, pretreatment specimens, critical for evaluation of biologic markers that could be affected by chemotherapy or radiation therapy, can be obtained. Furthermore, case–control studies can generally collect larger quantities and types of biologic samples and process them in ways (i.e., cryopreservation of lymphocytes and Epstein-Barr virus transformation to ensure large quantities of DNA) that are more sophisticated than is feasible in cohort studies. This offers the potential for conducting functional assays such as mutagen sensitivity *(62)*, which in general are not methodologically feasible in cohort settings.

Single Disease

Case–control studies are generally limited to one disease outcome but are less constrained by the rarity of the disease, whereas cohort studies (including full cohort, nested case–control, or case–cohort studies) may identify multiple disease end points. The focus of a case–control study on one disease entity permits greater time and resources to be devoted to the collection and documentation of disease information (i.e., pathology, staging, and tissue), providing a richer information base for study with the benefits of a potential reduction in misclassification and a deeper analysis. Obtaining disease-related data in cohorts entails mounting an effort that is generally less efficient and more costly. The advantage of cohort studies' ability to examine multiple outcomes may be somewhat limited by resources and logistics, limited exposure information, the diverse approaches to documenting disease incidence or mortality, and the rarity of some outcomes.

Costs for a series of case–control studies of different cancer sites can sometimes be reduced by sharing a single control group. When different diseases require different exposures, the partial questionnaire design may offer a reduction in the burden to respondents, thereby potentially increasing participation *(63)*. Even if these options are not feasible, using the same infrastructure for control selection for repeated studies can reduce costs.

Resources and Infrastructure

The vastly greater size of cohorts large enough to generate a substantial sample of diseased subjects and the time period required for the cohort to mature mean that a substantially greater initial investment is required to establish the cohort. For cohort studies that incorporate biologic materials, the infrastructure to support biospecimens' databases, freezers, and processing requires a correspondingly greater effort and cost. While all studies with biospecimens must consider the risk of untoward events (i.e., freezer failure), the anticipated long useful life of the samples from cohorts requires special emphasis on quality control and security issues (i.e., backup generators, monitoring, distributing samples among different freezers, etc.). In the next few years, however, the "cost per case" for studies fielded from a cohort will offer economies in comparison to fielding a new case–control study *(64)*.

Future Trends

The primacy of exposure in cancer etiology and the recognition that cancer is manifestly hereditary in certain families, but consistently a genetic disease on the cellular level mandate an interdisciplinary approach to elucidating their joint role in cancer. Family studies have proven critical to identifying high-penetrance genes. Mechanistic work and animal work have provided extraordinary insight into our fundamental understanding of cancer. Unraveling the complex origin of cancer that results in the vast burden of cancer mortality will require population-based studies that can uniquely determine the public heath impact of a putative risk factor, genetic or otherwise.

Population studies that include tissue collection and exposure assessment will allow opportunities for diverse hypotheses testing involving, e.g., relations between germline and tumor mutations, exposure relationships in questionnaire-based and molecular approaches, and exploitation of the developing human genome map for both population and family-based studies. There will be opportunities to explore for unique and newly hypothesized etiologic agents in stored material, to correlate gene and exposure markers with tumor mutations, and generally to explore mechanistic hypotheses. The next generation of studies that will strive to understand the interplay between environmental and genetic factors for cancer risk in the general population will be substantially larger than previous biologically based case–control studies in order to have adequate power to detect interactions between environmental and genetic factors *(38)*. Generally, larger study sizes should also permit new approaches requiring pooled DNA samples that exploit the increasingly detailed genetic map to conduct gene searches, i.e., linkage disequilibrium mapping *(65)* with the use of unrelated control subjects. Less invasive methods of genomic DNA collection, such as buccal cell collections *(61)*, will find increasing application and will benefit all study approaches involving biospecimens.

In the future, both case–control and cohort studies will have crucial and complementary roles in the investigation of genetic factors and their interaction with environmental exposures in the general population. As the number of cases with common tumors accumulate in existing cohorts over the coming years, nested case–control studies will have an increasingly important role, with special advantages deriving from large size, multiple outcomes (both cancer and non-cancer), and prospective expo-

sure assessment. Traditional case–control studies and their variants will be central for focused investigations of single cancers, where detailed exposure and intensive biologic sample collection are deemed necessary to integrated investigations. The argument for the case–control method will be particularly compelling when new or changing exposures must be quickly evaluated, when populations underrepresented in cohort studies must be rapidly investigated, or when less common tumors are the focus of study.

References

(1) Thomas DB, Karagas MR. Migrant studies. In: Schottenfeld D, Fraumeni JF Jr. Cancer epidemiology and prevention. 2nd ed. Oxford (U.K.): Oxford University Press, 1996. p. 236–54.

(2) Schottenfeld D, Fraumeni JF Jr. Cancer epidemiology and prevention. 2nd ed. Oxford (U.K.): Oxford University Press; 1996.

(3) Frame LT, Ambrosone CB, Kadlubar FF, Lang NP. Host–environment interactions that affect variability in human cancer susceptibility. In: Neumann DA, Kimmel CA, editors. Human variability in response to chemical exposures. Washington (DC): International Life Sciences Institute; 1998. p. 165-204.

(4) Lang M, Pelkonen O. Metabolism of xenobiotics and chemical carcinogenesis. In: Metabolic polymorphisms and susceptibility to cancer. IARC Sci Publ 1999;148:13–22.

(5) Lindor NM, Greene MH, the Mayo Familial Cancer Program. The concise handbook of family cancer syndromes. J Natl Cancer Inst 1998;90: 1039–71.

(6) Langholz B, Rothman N, Wacholder S, Thomas DC. Cohort studies for characterizing measured genes. Monogr Natl Cancer Inst 1999;26:39–42.

(7) Hill AB. The environment and disease: association or causation. Proc R Soc Med 1965;58:295–300.

(8) Rothman KJ. Modern epidemiology. 2nd ed. Boston (MA): Little Brown & Co.; 1998.

(9) Kukull WA, Schellenberg GD, Bowen JD, McCormick WC, Yu CE, Teri L, et al. Apolipoprotein E in Alzheimer's disease risk and case detection: a case–control study. J Clin Epidemiol 1996;49:1143–8

(10) Sanders AM, Strittmatter WJ, Schmechel D, George-Hyslop S, Pericak-Vance MA, Joo SH, et al. Association of apolipoprotein E allele e4 with late onset familial and sporadic Alzheimer's disease. Neurology 1993;43: 1467–72.

(11) Aird I, Bentall HH. A relationship between cancer of the stomach and the ABO blood groups. Br Med J 1953;1:799–801.

(12) Kellermann G, Shaw CR, Luyten-Kellermann M. Aryl hydrocarbon hydroxylase inducibility and bronchogenic carcinoma. N Engl J Med 1973; 289:934–7.

(13) Ayesh R, Idle JR, Richie JC, Crothers MJ, Hetzel MR. Metabolic oxidation phenotypes as markers for susceptibility to lung cancer. Nature 1984;312: 169–70.

(14) Seidegard J, Pero RW, Miller D, Beattie EJ. A glutathione transferase in human leukocytes as a marker for susceptibility to lung cancer. Carcinogenesis 1986;7:751–3.

(15) Lower GM, Nilsson T, Nelson CE, Wolf H, Gamsky TE, Bryan GT. *N*-Acetyltransferase phenotype and risk in urinary bladder cancer: approaches in molecular epidemiology. Preliminary results in Sweden and Denmark. Environ Health Perspect 1979;29:71–9.

(16) Nebert DW, Ingelman-Sundberg M, Daly AK. Genetic epidemiology of environmental toxicity and cancer susceptibility: human allelic polymorphisms in drug metabolizing genes, their functional importance, and nomenclature issues. Drug Metab Rev 1999;31:467–82.

(17) Caporaso N, Goldstein A. Issues involving biomarkers in the study of the genetics of human cancer. IARC Sci Publ 1997;142:237–50.

(18) Marcus PM, Vineis P, Rothman N. NAT2 slow acetylation and bladder cancer risk: a meta-analysis of 22 case–control studies conducted in the general population. Pharmacogenetics. In press 2000.

(19) Rebbeck TR. Molecular epidemiology of the human glutathione *S*-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. Cancer Epidemiol Biomarkers Prev 1997;6:733–43.

(20) Houlston RS. Glutathione *S*-transferase M1 status and lung cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 1999;8:675–83.

(21) D'Errico A, Taioli E, Chen X, Vineis P. Genetic metabolic polymorphisms and the risk of cancer: a review of the literature. Biomarkers 1996;1: 149–73.

(22) Kawajiri K. CYP1A1. IARC Sci Publ 1999;148:159–72.

(23) Wu XF, Amos CI, Kemp BL, Shi HH, Jiang H, Wan Y, et al. Cytochrome P450 2E1 *Dra*I polymorphisms in lung cancer in minority populations. Cancer Epidemiol Biomarkers Prev 1998;7:13–8.

(24) Shaw GL, Falk JN, Weiffenbach B, Nesbitt JC, Pass HI, Caporaso NE, et al. Genetic polymorphism of CYP2D6 and lung cancer risk. Cancer Epidemiol Biomarkers Prev 1998;7:215–9.

(25) Nazar-Stewart V, Vaughan TL, Burt RD, Chen C, Berwick M, Swanson GM. Glutathione *S*-transferase M1 and susceptibility to nasopharyngeal carcinoma. Cancer Epidemiol Biomarkers Prev 1999;8:547–51.

(26) Hildesheim A, Anderson LM, Chen CJ, Cheng YJ, Brinton LA, Daly AK, et al. CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. J Natl Cancer Inst 1997;89:1207–12.

(27) Harty LC, Caporaso NE, Hayes, RB, Winn DM, Bravo-Otero E, Blot WJ, et al. Alcohol dehydrogenase 3 genotype and risk of oral cavity and pharyngeal cancers. J Natl Cancer Inst 1997;89:1698–705.

(28) McGlynn KA, Rosvold EA, Lustbader ED, Hu Y, Clapper ML, Zhou T, et al. Susceptibility to hepatocellular carcinoma is associated with genetic variation in the enzymatic detoxification of aflatoxin B1. Proc Natl Acad Sci U S A 1995;92:2384–7.

(29) Lin DX, Tang YM, Peng Q, Lu SX, Ambrosone CB, Kadlubar FF. Susceptibility to esophageal cancer and genetic polymorphisms in glutathione *S*-transferases T1, P1, and M1 and cytochrome P450 2E1. Cancer Biomarkers Epidemiol Prev 1998;7:1013–8.

(30) Marcus PM, Hayes RB, Vineis P, Garcia-Closas M, Caporaso NE, Rothman N. NAT2 slow acetylation and cigarette smoking; a meta-analysis of their interactive effect on bladder cancer risk among 1,986 cases [abstract]. Proc Am Assoc Cancer Res 1999;40:212.

(31) Ambrosone CB, Freudenheim JL, Graham S, Marshall JR, Vena JE, Brasure JR, et al. Cigarette smoking, *N*-acetyltransferase 2 genetic polymorphism and breast cancer risk. JAMA 1996;276:1494–501.

(32) Kampman E, Slattery ML, Bigler J, Leppert M, Samowitz W, Caan BJ, et al. Meat consumption, genetic susceptibility, and colon cancer risk: a United States multicenter case–control study. Cancer Epidemiol Biomarkers Prev 1999;8:15–24.

(33) Irvine RA, Yu MC, Ross RK, Coetze GA. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. Cancer Res 1995;55:1937–40.

(34) Makridakis NM, Ross RK, Pike MC, Crocitto LE, Kolonel LN, Pearce CL, et al. Association of mis-sense substitution in SRD5A2 gene with prostate cancer in African-American and Hispanic men in Los Angeles, USA. Lancet 1999;354:975–8.

(35) Dean M, Jacobson LP, McFarlane G, Margolick JD, Jenkins FJ, Howard OM, et al. Reduced risk of AIDS lymphoma in individuals heterozygous for the CCR5-delta 32 mutation. Cancer Res 1999;59:3561–4.

(36) Longuemaux S, Delomenie C, Gallou C, Mejean A, Vincent-Viry M, Bouvier R, et al. Candidate genetic modifiers of individual susceptibility to renal cell carcinoma: a study of polymorphic human xenobiotic-metabolizing enzymes. Cancer Res 1999;59:2903–8.

(37) Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. BMJ 1952;2:1271–86.

(38) Rothman N, Garcia-Closas M, Stewart WJ, Lubin J. Impact of misclassification in studies of gene–environment interactions. IARC Sci Publ 1999; 148:89–96.

(39) Pearce N, Boffetta P. General issues of study design and analysis in the use of biomarkers in cancer epidemiology. IARC Sci Publ 1997;142:47–58.

(40) Landi MT, Caporaso N. Issues in the collection, processing and storage of biospecimens. IARC Sci Publ 1997;142:223–6.

(41) Schulte PA, Perera FP. Transitional studies. IARC Sci Publ 1997;142: 19–29.

(42) Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 1988;43:5205–26.

(43) Altshuler D, Kruglyak L, Lander E. Genetic polymorphism and disease. N Engl J Med 1998;338:1626.

(44) Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994; 265:2037–48.

(45) Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case–control studies of candidate genes and gene–environment interactions: basic family designs. Am J Epidemiol 1999;149:693–705.
(46) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Monogr Natl Cancer Inst 1999;26:31–7.
(47) Rothman N, Stewart W, Schulte PA. Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. Cancer Epidemiol Biomarkers Prev 1995;4:301–11.
(48) Khoury MJ, Beaty TH. Applications of the case–control method in genetic epidemiology. Epidemiol Rev 1994;16:134–50.
(49) Caporaso N, Goldstein A. Cancer genes: single and susceptibility: exposing the difference. Pharmacogenetics 1995;5:59–63.
(50) Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7.
(51) Schork NJ, Cardon LR, Xu X. The future of genetic epidemiology. Trends Genet 1998;14:266–72.
(52) Garcia-Closas M, Lubin J. Power and sample size calculations in case–control studies of gene–environment interactions. Am J Epidemiol 1999; 149:689–92.
(53) Garcia-Closas M, Rothman N, Lubin J. Misclassification in case–control studies of gene–environment interactions: assessment of bias and sample size. Cancer Epidemiol Biomarkers Prev. In press 1999.
(54) Khoury MJ. Genetic epidemiology and the future of disease prevention and public health. Epidemiol Rev 1997;19:175–80.
(55) Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population case–control studies. Stat Med 1994;13:153–62.
(56) Yang Q, Khoury MJ. Evolving methods in genetic epidemiology. Gene–environment interaction in epidemiological research. Epidemiol Rev 1997; 19:33–43.
(57) Khoury MJ, Yang Q. The future of genetic studies of complex human diseases: an epidemiologic perspective. Epidemiology 1998;9:350–4.
(58) Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case–control studies. Am J Epidemiol 1992;135:1019–50.
(59) Breslow NE, Day NE. Statistical methods in cancer research, vol 1. The analysis of case–control studies. IARC Sci Publ 1980;32:84–119.
(60) Hunter DJ. Methodological issues in the use of biological markers in cancer epidemiology: cohort studies. IARC Sci Publ 1997;142:39–46.
(61) Lum A, Le Marchand L. A simple mouthwash method for obtaining genomic DNA in molecular epidemiologic studies. Cancer Epidemiol Biomarkers Prev 1998;7:719–24.
(62) Wu X, Gu J, Hong WK, Lee JJ, Amos CI, Jiang H, et al. Benzo[*a*]pyrene diol epoxide and bleomycin sensitivity to cancer of upper aerodigestive tract. J Natl Cancer Inst 1998;90:1393–9.
(63) Wacholder S, Carroll RJ, Pee D, Gail MH. The partial design for case–control studies. Stat Med 1994;13:623–34.
(64) Potter JD. Logistics and design issues in the use of biological samples in observational epidemiology. IARC Sci Publ 1997;142:31–7.
(65) Barcellos LF, Klitz W, Field LL, Tobias R, Bowcock AM, Wilson R, et al. Association mapping of disease loci, by use of a pooled DNA genomic screen. Am J Hum Genet 1997;61:734–47.

Note

We thank Alisa Goldstein for thoughtful review of this work.

Family-Based Association Studies

W. James Gauderman, John S. Witte, Duncan C. Thomas

We review case–control designs for studying gene associations in which relatives of case patients are used as control subjects. These designs have the advantage that they avoid the problem of population stratification that can lead to spurious associations with noncausal genes. We focus on designs that use sibling, cousin, or pseudosibling controls, the latter formed as the set of genotypes not transmitted to the case from his or her parents. We describe a common conditional likelihood framework for use in analyzing data from any of these designs and review what is known about the validity of the various design and analysis combinations for estimating the genetic relative risk. We also present comparisons of efficiency for each of the family-based designs relative to the standard population-control design in which unrelated controls are selected from the source population of cases. Because of overmatching on genotype, the use of sibling controls leads to estimates of genetic relative risk that are approximately half as efficient as those obtained with the use of population controls, while relative efficiency for cousin controls is approximately 90%. However, we find that, for a rare gene, the sibling-control design can lead to improved efficiency for estimating a $G \times E$ interaction effect. We also review some restricted designs that can substantially improve efficiency, e.g., restriction of the sample to case-sibling pairs with an affected parent. We conclude that family-based case–control studies are an attractive alternative to population-based case–control designs using unrelated control subjects. [Monogr Natl Cancer Inst 1999;26:31–7]

Association studies are routinely used by epidemiologists to investigate the relationship between an exposure and a disease. With the recent increase in the availability of genetic information, these exposures may now include genotypes at one or more susceptibility, candidate, or marker loci. The goals of genetic association studies will differ, depending on the state of knowledge about the given disease. For example, once a susceptibility locus has been cloned (e.g., BRCA1 for breast cancer), the goals include estimating the relative risk (RR) and penetrance associated with specific mutations and testing for interaction with environmental exposures or other genes *(1)*. If a candidate locus has been identified (e.g., the androgen receptor for prostate cancer), the primary goal is testing the null hypothesis of no association between the locus and the disease. Finally, if little is known about specific loci for the disease (e.g., multiple sclerosis), multiple tests of association with finely spaced markers may be used to screen the genome for candidate regions in the hopes of detecting linkage disequilibrium with markers close to one or more disease loci.

The case–control design is generally considered the design of choice for studying rare diseases, although suitably designed cohort studies, particularly family-based cohort studies *(2)*, are also useful in some circumstances. For results to be generalizable, the selection of case patients in a case–control study should be population based. This process is relatively straightforward for diseases like cancer, for which population-based registries are available, provided one can identify cases rapidly enough to enroll them and obtain blood samples. For effect estimates and hypothesis tests to be valid, control subjects should be selected from the same source population as the cases. In the situation of disorders with a genetic basis, this implies that cases and controls should derive from a similar genetic background.

One approach used to satisfy this requirement is to match cases and controls on their race or ethnicity. However, even within subgroups, strong variation can be found in allele frequencies at many genetic loci (e.g., the gradient in human leukocyte antigen allele frequencies from northern to southern Europeans). An additional complication is that, in many places, a given subject may represent a mixture of genetic backgrounds as a result of intermarriage between ancestors of varied ethnicities, and, as a practical issue, many subjects will not know with certainty their complete ancestral background. This uncertainty makes finding a suitable population-based control for such subjects very difficult. If the allele frequency at a particular genetic locus varies across ethnic groups and if ethnicity (or some unobserved factor that varies by ethnicity) is a risk factor for disease independent of that locus, then failure to adequately control for ethnicity can result in false associations between the gene and the disease *(3–5)*. This phenomenon is often referred to as population stratification by geneticists and as confounding by epidemiologists. The unobserved ethnic factor associated with disease can be either another gene or an environmental factor. An example of such confounding is the reported association between the Gm locus and non-insulin-dependent diabetes mellitus (NIDDM) in American Indians that disappeared when the analysis was restricted to full-heritage Pima-Papago Indians *(6)*. The likely explanation for this finding was that the Gm locus served as a surrogate for Caucasian heritage and that the risk of NIDDM varied with this level of ancestry.

Recently, much interest has been focused on the use of family-based controls to avoid the problem of ethnic confounding. One approach is to match each case with one or more unaffected siblings *(7,8)* or cousins *(8)* and to use analytic techniques for matched case–control studies *(9)* to estimate effects and to test hypotheses. A second approach is to match each case to a set of "pseudosiblings," formed as the set of genotypes that was not transmitted from the parents to the case. Several methods have been proposed for testing candidate gene associations and for estimating genetic RRs, including the transmission disequilibrium test (TDT), conditional logistic regression, and haplotype-sharing techniques *(4,10–19)*. Both the sib-control and pseudo-

Affiliations of authors: W. J. Gauderman, D. C. Thomas, Department of Preventive Medicine, University of Southern California, Los Angeles; J. S. Witte, Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Correspondence to: W. James Gauderman, Ph.D., Department of Preventive Medicine, University of Southern California, 1540 Alcazar St., Suite 220, Los Angeles, CA 90089 (e-mail: jimg@rcf.usc.edu).

See "Note" following "References."

© Oxford University Press

sib-control approaches have the advantage that they provide perfect matching on ethnicity.

We review the basic family-based case–control designs, describe a general approach to examine their validity and efficiency, and summarize what is known about their relative efficiency for estimating the genetic RR. We also describe variations on currently proposed designs that may be useful in some circumstances.

Designs

For a disease like cancer with variable age of onset, we consider the genetic RR parameter of interest to be the ratio of age-specific incidence rates (i.e., the hazard rate ratio). With this choice, the odds ratio from any matched case–control design is a consistent estimator of the RR, provided controls are randomly selected from the "risk set" comprising those members of the population at risk who were disease free at the age at which the case was affected. Indeed, exclusion of subjects who later developed the disease of interest will bias odds ratio estimates away from the null *(20)*. There is then no need for a rare disease assumption *(21)*. Control subjects should also be matched to case patients on any potential confounders and generally should be matched on sex (particularly for sex-specific diseases).

Sib Controls

Instead of defining the source population as the entire population, one could consider only the immediate or more distant family members of the case as potential controls, leading to the various designs considered here. For example, in the sib-matched case–control design, the investigator matches each case patient to one or more unaffected sibling controls. The principles of risk-set sampling require that controls have attained the age of the case and still be disease free. If only recently incident cases are included, this criteria essentially restricts control selection to older siblings. Of course, a sibling who is younger than the case may achieve the case's age of diagnosis during the study period and then become eligible as a control. Use of siblings who have not yet attained the age of the case may lead to effect estimates that are biased away from the null, but this bias could theoretically be corrected with the use of knowledge of the population rates. Inclusion of such siblings would also pose problems if time-dependent covariates are involved.

Although genotypes do not change with age, a restriction to younger sibs could lead to confounding of the effects of environmental exposures that have secular trends (e.g., oral contraceptive use) and conceivably confounding of the effects of any genotypes with which such risk factors were associated. As in any case–control study of time-dependent factors, the exposure status of cases and controls should be compared at a common "reference age," such as the case's age at diagnosis (or some common interval prior to it to allow for latency). In addition to being perfectly matched on ethnicity, siblings will also be matched on many other potential confounding variables. Although this match offers protection from bias, siblings are likely to be overmatched on many factors (including genotype) that will result in less efficient parameter estimation. This situation will be explored quantitatively below.

From a practical standpoint, the use of sibling controls may offer several nonstatistical advantages over population controls. The occurrence of disease in the case may make his or her relatives much more willing to participate than an unrelated subject from the general population. In addition to reducing cost, this willingness may result in the control being more careful in filling out a risk-factor questionnaire. Because the case and sibling will share many exposures, researchers will be able to cross-validate questionnaire information that has been obtained from the case and control, such as the types of cancer in their ancestors, or to ask comparative questions, such as which of the two was more exposed to particular environmental factors. Many groups have, or are in the process of collecting, family-based cancer data resources. For example, the Cancer Surveillance Project for Orange and San Diego Counties routinely abstracts family history information on first- and second-degree relatives and first cousins of all cases; this resource is the basis of a population-based family study of breast and ovarian cancers involving a family-history stratified sample of cases *(22)*. Once a resource such as this project has been established, selection of sibling controls can be much less expensive than finding controls from the general population. Conversely, not all cases will have an eligible and willing sibling available; in addition to the obvious loss of sample size, it is possible that selection bias could arise if availability of a sib control were related both to disease risk and to allele frequency.

Cousin Controls

Instead of a sibling, one could obtain another relative of the case as a control. If one is also studying risk factors of which distribution varies by generation, controls should probably be drawn from the same generation, such as first cousins. Compared with a sibling control, the advantage of a cousin is that one may be able to obtain closer matching on age and year of birth, with less loss in efficiency because the case and cousin are not as closely matched on genotype. The trade-off is that there is no longer the absolute protection from ethnic confounding because the case and cousin have only one side of their families in common and there is no guarantee that the two unrelated parents of the case and cousin derive from the same ethnic background. In this circumstance, one might want to select two cousin controls, one from each side of the family, but it remains to be shown that this will provide adequate protection from bias. From a practical standpoint, cousin controls have many of the same advantages as sibling controls, including increased willingness to participate and possible pre-identification through a family-based data resource. As in the sibling design, not all cases will have an eligible and willing cousin control, but there is generally a larger pool from which to choose.

Pseudosibling Controls

In this design, no actual controls are selected. Instead, genotypic data are obtained on the parents of the case, and the genotype transmitted to the case is then compared with the three genotypes (pseudosiblings) that were not transmitted to the case. Suppose we label the alleles of the two parents (a,b) and (c,d) and the case's alleles as (a,c) (recognizing that some of these alleles may be identical by state). Then the three pseudosibling genotypes are (a,d), (b,c), and (b,d) and the question that this design seeks to address is whether a specific allele or genotype occurs more commonly in cases than in their pseudosibs. Conditional logistic regression for 1:3 matched case–control studies is the appropriate analysis for such data *(11)*. The TDT, which simply compares the case with his or her "antisib" (b,d), has been shown to be the score test from the conditional logistic

likelihood under a multiplicative model for dominance, in which the homozygote RR is the square of the heterozygote RR *(4,16,19)*.

Both the sib-matched case–control and the pseudosib (or TDT) designs test the same null hypothesis, i.e., that there is no association or no linkage between the candidate gene (or marker) and the underlying trait gene. Thus, neither design will detect association with a gene that is in disequilibrium with a causal gene (e.g., because of population stratification) but that is not linked to that causal gene. Essentially, sib controls can be thought of as a finite realization of genotypes from the theoretical distribution of pseudosib genotypes, with the only fundamental difference being that real sibs are required to have survived to the age of the case. The lack of this restriction in the pseudosib design produces an estimator of the genetic RR that is biased toward the null by an amount that disappears with increasing disease rarity, although the validity of the hypothesis test is not affected *(8)*.

As with the sibling- and cousin-control designs, parents are more likely to be willing to participate than a population control, and the design will take advantage of existing information available in a family-based data resource. In practice, the utility of this design is limited to disorders that occur at young enough ages that parents of the case are still likely to be alive. This limitation excludes many cancers, unless the focus is on younger onset cases. It has been shown that, if the genotype is missing on one parent, there can be bias in the TDT *(23)*. An alternative approach when parental data are missing is to use the sib-TDT *(24)*, which involves a comparison of the genotype of an affected sibling to that in an unaffected sibling, and is similar to the sibling-control approach described above. One can also combine the TDT and sib-TDT, using parental genotypes if they are available and siblings if they are not *(24)*. However, if there are multiple affected subjects (or multiple unaffected siblings for the sib-TDT), the TDT and sib-TDT provide only a valid test of linkage; the test of association will have an inflated type I error rate.

Restricted Designs

For diseases that are not too rare, one might consider any of the above designs with an additional restriction to subjects with a positive family history. The rationale would be to increase the allele frequency in the sample, thereby improving the statistical efficiency for detecting associations with rare genes. However, care must be taken that any restriction applied to cases is applied equally to controls. For example, if one required the case to have an affected first-degree relative, one would have to make the same requirement for controls. For a design with population controls, this requirement might entail some form of multistage sampling *(25,26)*, in which one obtains family history information on an unrestricted series of potential cases and controls and then selects a subsample of those with a positive history. Sib controls are automatically matched on family history (among sibs, parents, and more distant relatives, but not their offspring), and such sibships might be easily identified from registries that contain family history data. Cousin controls with comparable family histories are not as easily identified, although case–cousin pairs that share an affected grandparent would be a valid comparison, as would those that each have an affected sibling. However, case–cousin pairs that each have an affected parent would be a valid comparison only if the two parents were related to each other or if neither parent was a relative of the other pair member. Such case–cousin pairs with two affected relatives would generally be quite uncommon.

Comparison of Designs

We now describe our basic approach to comparing the validity and relative efficiency for estimating the genetic RR for various family-control and population-control designs.

We assume that the data consist of diseased subjects (cases) and one or more matched controls (real or pseudosiblings) for each case. Let d_{ij} denote the disease status of subject j in matched set i, and let g_{ij} denote the genotype at some locus of interest. For simplicity, we assume that the alleles at the locus can be classified as either mutant (denoted by A) or normal (denoted by a), with population frequency q of the A allele, although the methods are easily extended to genes with more than two alleles. Let $G(g)$ denote a genetic covariate with values $G(g) = 0$ when $g = aa$, $G(g) = 1$ when $g = AA$, and $G(g) = \Delta$ when $g = Aa$. The parameter Δ is coded to reflect an assumed mode of inheritance, with $\Delta = 1$ corresponding to dominant inheritance, $\Delta = 0$ to recessive inheritance, and $\Delta = 0.5$ to multiplicative (or log-additive) inheritance; this parameter can also be estimated in a general codominant model.

For a binary trait, we assume a logistic model for penetrance, i.e.,

$$\text{logit}[Pr(d = 1|g)] = \alpha_i + \beta G(g), \qquad [1]$$

where α_i denotes the logit of the baseline risk for noncarriers in matched set i, and β is the log-RR for carriers of a mutation. For matched pairs data, the conditional likelihood is a function of only β, which we assume is common across matched pairs. If β were variable across the population, then a study would estimate some form of weighted average of the distribution of β values, the particular weighting being somewhat different for population-control versus family-control designs. In the family-control designs, we assume that the disease outcomes among relatives are conditionally independent, given their genotypes. Letting g_{i1} denote the genotype of the case in the ith matched pair, the conditional logistic likelihood is

$$L(\beta) = \prod_i \frac{\exp(\beta G(g_{i1}))}{\sum_{j \in M_i} \exp(\beta G(g_{ij}))}, \qquad [2]$$

where M_i denotes the set of subjects in the ith case–control set. For the case–pseudosib design, j ranges over the case and the three pseudosibs.

For a disease of variable age at onset, essentially the same likelihood can be derived from Cox's proportional hazards model,

$$\lambda(t,g) = \lambda_0(t) \exp(\beta G(g)),$$

where $\lambda(t,g)$ denotes the genotype-specific incidence rate at age t and $\lambda_0(t)$ denotes an unspecified set of baseline rates in noncarriers. Equation 2 then results when the controls are drawn at random from the risk set for the ith case.

The models above can be expanded to include one or more environmental covariates (z) and gene–environment interaction terms. In this case, the logistic model becomes

$$\text{logit}[Pr(d = 1|g,z)] = \alpha_i + \beta G(g) + \gamma z + \eta G(g) z,$$

with an analogous extension to Cox's proportional hazards model. For either model, the conditional likelihood is

$$L(\beta, \gamma, \eta) = \prod_i \frac{\exp[\beta G(g_{i1}) + \gamma z_{i1} + \eta G(g_{i1}) z_{i1}]}{\sum_{j \in M_i} \exp[\beta G(g_{ij}) + \gamma z_{ij} + \eta G(g_{ij}) z_{ij}]}.$$

In the pseudosibling design, z_{ij} is set equal to z_{i1} for all j, precluding estimation of the environmental main effect parameter (γ) and requiring an assumption of independence of the genetic and environmental factors conditional on parental genotypes for valid estimation of η.

To assess the validity of a design or analysis combination for estimation of β, we computed the expectation of the score statistic (the first derivative with respect to β of the log likelihood) under the true model. If this expectation is zero, then the estimator is said to be Fisher consistent, meaning that the maximum likelihood estimate will converge to the true value with increasing sample size. In this case, the asymptotic relative efficiency (ARE) for estimating β for one design compared with another is defined as the inverse of the ratio of their expected variances of $\hat{\beta}$ under the alternative hypothesis or equivalently as the ratio of the sample size required to attain the same precision and power. We compute the expected variance of $\hat{\beta}$ as the inverse of the Fisher information, evaluated at the true value of the parameters (α_0, β_0, q_0). For comparability across several parameter values, we fixed the population prevalence of the disease (K_p) and the attributable risk (AR) and then, for given values of the log-RR (β_0), solved the following two equations for α_0 and q_0:

$$AR = \frac{\sum_g (e^{\beta_0 G(g)} - 1)\, Pr(g|q_0)}{\sum_g (e^{\beta_0 G(g)})\, Pr(g|q_0)}$$

and

$$K_p = \sum_g Pr(d = 1|G(g), \alpha_0, \beta_0)\, Pr(g|q_0).$$

The factor $Pr(g|q_0)$ was computed assuming Hardy–Weinberg equilibrium, and the penetrance factor in the equation for K_p was computed as the anti-logit of the expression in equation 1. Letting *Rel* denote the relationship between the case and the control and assuming a 1 : 1 matched design, the Fisher information was computed as

$$E[I(\beta|Rel)] = \sum_g I(\beta|g)\, Pr(g|d_1 = 1, d_2 = 0, \alpha_0, \beta_0, q_0, Rel)$$

$$= \frac{\sum_g I(\beta|g)\, Pr(d_1 = 1|g_1)\, Pr(d_2 = 0|g_2)\, Pr(g|Rel, q_0)}{\sum_g Pr(d_1 = 1|g_1)\, Pr(d_2 = 0|g_2)\, Pr(g|Rel, q_0)}, \quad [3]$$

where $g = (g_1, g_2)$ and $I(\beta)$ is the observed information, i.e. the negative of the matrix of second partial derivatives of the conditional log-likelihood. One can see from equation 3 that the joint distribution of the case and control genotypes is the factor that differentiates the informativeness of the various designs. If the case and control are unrelated, $Pr(g|Rel, q_0) = Pr(g_1|q_0)Pr(g_2|q_0)$, and the weight is determined solely by the allele frequency. However, if the case and control are siblings, $Pr(g|Rel, q_0) = \sum_{g_f, g_m} Pr(g_1|g_f, g_m)Pr(g_2|g_f, g_m)Pr(g_f|q_0)Pr(g_m|q_0)$ and the weight is a function of both the allele frequency and the genetic relationship between the pair. Note that, although computation of the expected information for the sib-matched design involves a summation over parental genotypes (g_f, g_m), the actual information depends only on the joint distribution of the genotypes of the case and sibling control.

Results

In the presence of population stratification, the amount of bias in the estimate of genetic RR when using the population-based case–control design depends on the true RR and the ratio of allele frequencies in the strata *(8)*. The sib-control design is always consistent, the pseudosib design is approximately consistent for a rare disease but inconsistent for a common disease, and the consistency of the cousin design will depend on whether the unrelated parents of the cousins come from the same or different population strata. The bias in the pseudosib design for a common disease occurs even in the absence of population stratification, although a method has been proposed for correcting this bias *(8)*.

Fig. 1 provides a summary of the ARE of the three basic family designs for estimating β, relative to case–control studies using unrelated controls. The results are based on a disease with population prevalence K_p = 1% and AR = 5%, although the relative efficiencies are not substantially affected by these two parameters. Under a dominant model, the ARE is approximately 50% using sib controls, 88% using cousin controls, and 100% using pseudosib controls, regardless of the true underlying value of the genetic RR. Under a multiplicative model, these three AREs are nearly identical to those for the dominant model (data not shown). For the recessive model, the relative efficiencies are higher than for the dominant model in all three designs. As the genetic RR ranges from 2 to 20, the AREs range from 66% to 72% using sib controls, from 95% to 99% using cousin controls, and from 150% to 260% using pseudosib controls. Although on a per-case basis the pseudosib design is statistically more efficient than unrelated controls for a recessive gene, this design requires three genotypes per case rather than two and so may be less cost efficient if the cost of genotyping is high in relation to the cost of subject enrollment.

To provide some intuition to account for these efficiency comparisons, Table 1 provides the expected number of case–control pairs in the unrelated-control and sib-control designs for a dominant and a recessive gene and a particular choice of parameter values. No population stratification is assumed, and, in all the designs shown, the McNemar odds ratio (*c/b*) provides a good approximation to the assumed RR of 20. Compared with the unrelated-control design, use of the sibling-control design results in a larger proportion of genotype-concordant pairs (cells *a* and *d*) because of overmatching, and thus a smaller number of discordant pairs (*b* and *c*) on which the variance of $\hat{\beta}$ is determined. For relatively rare genes, it is evident that the primary determinant of this variance is the number *b* of case–noncarrier and control–carrier pairs.

Similar comparisons of relative efficiency for estimating the gene–environment interaction parameter η have also been carried out *(8)*. The efficiency for a particular design in this case depends on the distribution of the three types of discordant pairs: 1) genotype concordant and exposure discordant, 2) genotype discordant and exposure concordant, and 3) jointly discordant.

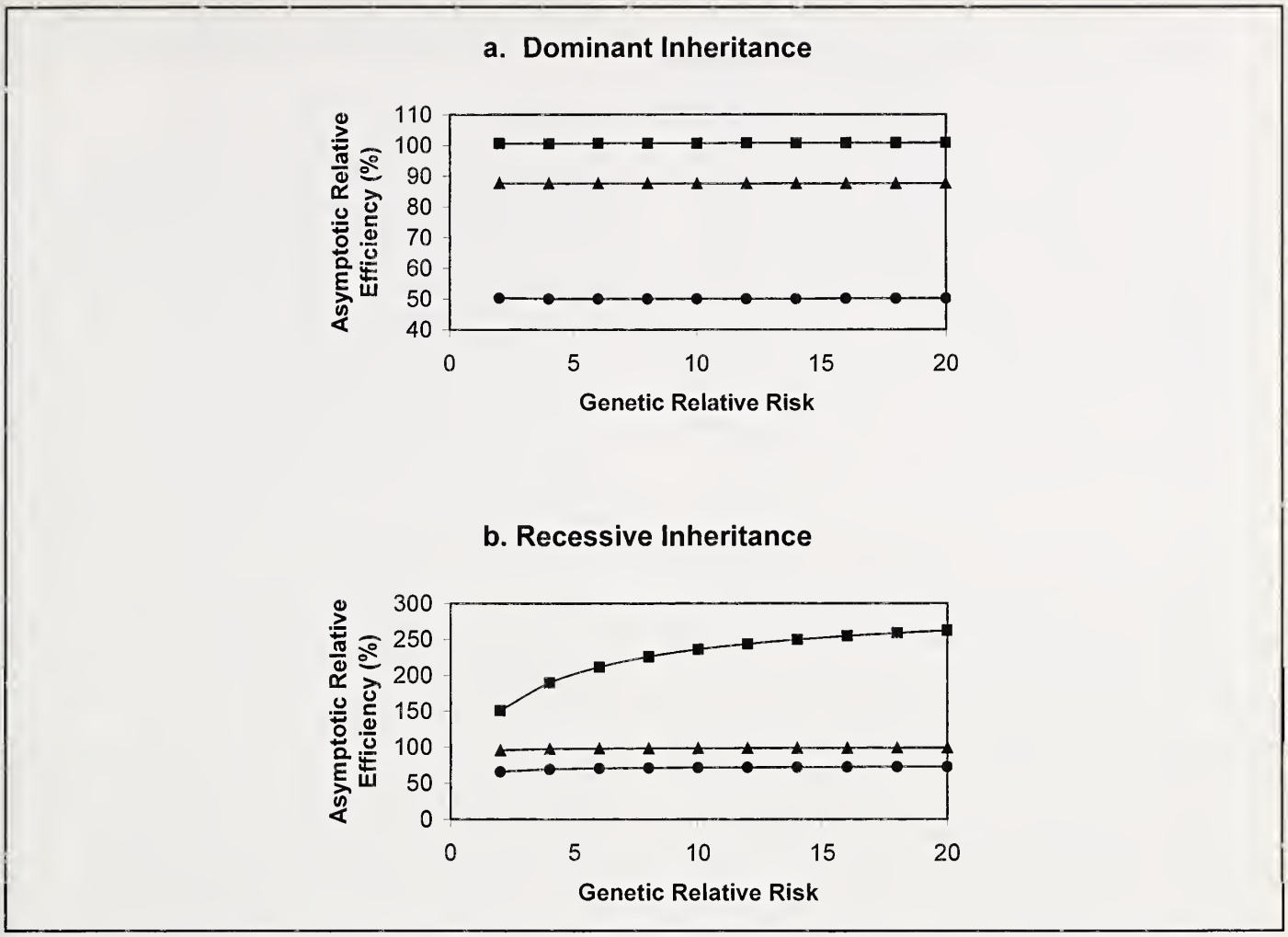


Fig. 1. Asymptotic relative efficiency versus genetic relative risk for the case–sibling (**circle**), case–cousin (**triangle**), and pseudosibling (**square**) designs, relative to the unrelated-control design, assuming disease prevalence of 1%, attributable risk of 5%, and either dominant (**a**) or recessive (**b**) inheritance.

Table 1. Illustrative example of the expected case–control genotype distributions for the unrelated-control and sib-control designs, under dominant and recessive models*

			Case genotype			
Model	Control type	Control genotype	Carrier	Noncarrier	$\mathrm{Var}(\hat{\beta}) = 1/b + 1/c$	ARE = V_1/V_2, %
Dominant	1) Unrelated	Carrier	$a = 46$	$b = 129$		
		Noncarrier	$c = 2576$	$d = 7249$	0.00814	
	2) Sibling	Carrier	$a = 1188$	$b = 66$		
		Noncarrier	$c = 1311$	$d = 7435$	0.01591	51
Recessive	1) Unrelated	Carrier	$a = 45$	$b = 127$		
		Noncarrier	$c = 2546$	$d = 7282$	0.00826	
	2) Sibling	Carrier	$a = 745$	$b = 88$		
		Noncarrier	$c = 1770$	$d = 7397$	0.01193	69

*Expected distributions were computed assuming the population disease prevalence = 1%, relative risk = 20, allele frequency = 0.14 (recessive) or 0.01 (dominant), and 10 000 case–control pairs in each design. ARE = asymptomatic relative efficiency.

For a rare gene, the use of sib controls can be substantially more efficient than the use of population controls for estimating a $G \times E$ effect. The reason is that, when the gene is rare, efficiency is determined primarily by the number of genotype concordant, exposure discordant pairs, although the other two types of discordant pairs also contribute (Table 2). Because of the partial matching on genotype, genotype concordant pairs are more common within sibships than within a case and unrelated control pair, leading to the improved efficiency. However, if sibs are also highly concordant for environmental exposure, this situation will tend to reduce their efficiency relative to unrelated controls.

In contrast to the basic designs, the relative efficiency of the restricted designs for estimating β depends strongly on the genetic RR and the AR but depends only weakly on the population prevalence of disease. Table 3 compares several restricted de-

Table 2. Illustrative example of the expected case–control genotype × environment distributions under a dominant model*

Genotype†	Environment‡	Unrelated control subjects	Sibling control subjects
C–C	E–U	155	418
C–C	U–E	207	1578
C–N	E–E	44	34
N–C	E–E	883	688
C–N	E–U	883	1124
N–C	U–E	66	30

*Expected distributions were computed assuming genetic relative risk = 10, environmental relative risk = 2, interaction relative risk = 2, allele frequency = 0.1, exposure prevalence = 0.25, and sibling exposure concordance odds ratio = 2.

†Genotype of the case–genotype of the control (C = carrier, N = noncarrier).

‡Exposure status in the case–exposure status in the control (E = exposed; U = unexposed).

signs for $K_p = 1\%$, several modes of inheritance, and a range of relative and ARs. Generally, the efficiency gains are greatest for genes with low AR and large RR, i.e., for rare major susceptibility genes. Across inheritance modes, efficiency gains in the restricted designs are greatest for a dominant gene. In the sibling-control design, restriction to pairs with an affected parent (design SAP) substantially improves efficiency for a dominant gene, whereas, if the restriction is to pairs with an additional affected sibling (design SAS), one can expect substantial efficiency gains for either a recessive or a dominant gene.

Absolute power can be computed with the use of standard methods once the expected distribution of case–control genotype probabilities has been computed. For example, using the data in Table 1, for a recessive gene with $q = 0.14$ and RR = 20, we would expect $c = 17.7\%$ and $b = 0.88\%$ of sib-matched case–control pairs to be informative, leading to a McNemar test of $\chi^2 = (c - b)^2/(c + b) = 0.152N$. To obtain 90% power at a two-sided 5% significance level, one would therefore require $N = (1.96 + 1.28)^2/0.152 = 69$ matched pairs. Of course, for smaller RRs, the required sample size would be larger. One can also use a standard software program to compute sample size for a case and unrelated-control design and then use the values plotted in Fig. 1 to adjust the sample size to a family-controlled design. For example, if one assumed a dominant model and the required number of pairs for a study with unrelated controls was 100, the necessary number of pairs would be 200 (100/0.5) for a case–sib study, or 114 (100/0.88) for a case–cousin study.

Discussion

We have argued that family-based case–control studies offer an attractive alternative to population-based case–control designs using unrelated controls. Their primary advantage is that they overcome the problem of population stratification that can lead to spurious associations with noncausal genes that are not even linked with any causal genes. The sibling and pseudosib designs completely avoid this problem, whereas the cousin-control design avoids it only approximately to the extent that families tend to marry within ethnic groups. This protection from bias is arguably worth the penalty of reduced statistical efficiency resulting from overmatching on genotype. We have also shown that on a per-case basis, the pseudosib design can be more efficient than the unrelated-control design and that restriction to multiple case families can lead to even more efficient designs, if done appropriately. Finally, we have argued that family-based designs offer certain nonstatistical advantages, such as improved cooperation and reduced cost, that must be weighed against the potential loss in sample size from cases who do have a suitable family control and the potential selection bias if such losses are nondifferential.

A spin-off of these family-based designs, particularly those involving cousin controls or restriction to multiple-case families, is the availability of phenotype information on other family members not involved themselves as cases or controls, whose genotypes may not be known. We have considered here the analysis of only the measured genotypes for the selected cases and their matched controls. To take advantage of the entire vector $\boldsymbol{d}$ of phenotype data on family members, one could conduct a "modified segregation analysis" in which one forms a likelihood by summing over all possible genotypes of the untyped individuals $\boldsymbol{g}_u$ conditional on the observed genotypes $\boldsymbol{g}_o$. The ascertainment process (Asc) (e.g., that each family contains at least one case and at least one unaffected sibling) can be addressed either by forming a "retrospective" likelihood from terms of the form $Pr(\boldsymbol{g}_o|\boldsymbol{d})$ or by modeling the ascertainment process explicitly in a "prospective" likelihood $Pr(\boldsymbol{d}|\boldsymbol{g}_o, Asc)$ or "joint" likelihood $Pr(\boldsymbol{d}, \boldsymbol{g}_o|Asc)$. For example the joint likelihood for a single family would be computed as

Table 3. Asymptotic relative efficiencies of various family-based case–control designs with restrictions on family history, relative to the unrestricted population-control design, assuming the population disease rate is 1%*

Design	Relative risk	Recessive		Multiplicative		Dominant	
		AR = 0.05	AR = 0.20	AR = 0.05	AR = 0.20	AR = 0.05	AR = 0.20
		ARE (%)		ARE (%)		ARE (%)	
SAP	2	71	58	58	53	70	61
	20	105	104	121	96	332	184
SAS	2	81	62	59	54	70	61
	20	313	197	124	103	332	184
CAG	2	102	96	94	90	102	93
	20	134	144	142	124	322	213
PAP	2	160	123	117	108	141	119
	20	318	273	248	204	654	351

*Design codes: SAP = siblings with affected parent; SAS = siblings with affected sibling; CAG = cousins with affected grandparent; PAP = pseudosiblings with affected parent; AR = attributable risk.

$$L(\alpha,\beta,q) = Pr(\boldsymbol{d},\boldsymbol{g}_o|Asc) = \frac{\sum_{g_u} Pr(Asc|\boldsymbol{d})\, Pr(\boldsymbol{d}|\boldsymbol{g}_o,\boldsymbol{g}_u,\alpha,\beta)\, Pr(\boldsymbol{g}_o,\boldsymbol{g}_u|q)}{\sum_{g_u}\sum_{d} Pr(Asc|\boldsymbol{d}) Pr(\boldsymbol{d}|\boldsymbol{g}_o,\boldsymbol{g}_u,\alpha,\beta)\, Pr(\boldsymbol{g}_o,\boldsymbol{g}_u|q)},$$

where the second sum in the denominator is taken over all possible vectors of disease status within the family. The likelihood for a set of families would be computed as the product of family-specific likelihood contributions.

An advantage of these segregation likelihoods is that they need not be restricted to families with at least one case and one unaffected relative. For example, if the initial ascertainment is based on selection of affected case patients from a population registry, all cases and their families can be included using the above likelihoods, while only those cases with an eligible unaffected sibling will be used in the conditional logistic likelihood for the case–sib design. However, whereas the conditional logistic likelihood depends only on the genetic RR parameter β, the segregation likelihoods also involve the baseline risk α and allele frequency q as nuisance parameters. Nevertheless, preliminary calculations indicate that incorporation of phenotypic data on relatives can lead to substantial gains in information compared with a case–control design, despite the need to estimate these additional parameters. Another disadvantage of the segregation likelihoods is the greater potential for bias if the form of the model is misspecified, e.g., if one were to assume the parameters were homogeneous across the population when in fact they were variable or if there was additional dependency within families that was not correctly modeled *(27)*.

An additional benefit of these family-based designs is that they can provide a resource for subsequent segregation and linkage analyses to test for and to localize additional genes, after accounting for any measured genes that may partially explain the observed familial aggregation *(28,29)*. To facilitate such studies, it would be helpful to have population-based disease registries with at least some family history data available, even if imperfect. This type of resource would more easily allow the ascertainment of cases with various types of family history, particularly the designs involving restriction to multiple-case families. In summary, family-based case–control designs have a number of attractive features that make them worth considering when designing a gene-association study for cancer or some other complex disease.

References

(1) Goldstein AM, Andrieu N. Detection of interaction involving identified genes: available study designs. Monogr Natl Cancer Inst 1999;26:49–54.

(2) Gail MH, Pee D, Benichou J, Carroll R. Designing studies to estimate the penetrance of an identified autosomal dominant mutation: cohort, case–control, and genotype–proband designs. Genet Epidemiol 1999;16:15–39.

(3) Khoury MJ, Beaty TH. Applications of the case–control method in genetic epidemiology. Epidemiol Rev 1994;16:134–50.

(4) Schaid DJ, Sommer SS. Comparison of statistics for candidate-gene association studies. Am J Hum Genet 1994;55:402–9.

(5) Caporaso N, Rothman N, Wacholder S. Case–control studies of common alleles and environmental factors. Monogr Natl Cancer Inst 1999;26:25–30.

(6) Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. $Gm^{3,5,13,14}$ and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 1988;43:520–6.

(7) Curtis D. Use of siblings as controls in case–control association studies. Ann Hum Genet 1997;61:319–33.

(8) Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case–control studies of candidate genes and gene–environment interactions: basic family designs. Am J Epidemiol 1999;149:693–705.

(9) Breslow NE, Day NE. Statistical methods in cancer research: I. The analysis of case–control studies. Vol. 32. Lyon (France): IARC Scientific Publications; 1989.

(10) Rubinstein P, Walker M, Carpenter C, Carrier J, Krassner J, Falk C, et al. Genetics of HLA disease association: the use of the haplotype relative risk (HRR) and the "Haplo-Delta" (Dh) estimates in juvenile diabetes from three racial groups. Hum Immunol 1981;3:384.

(11) Self SG, Longton G, Kopecky KJ, Liang KY. On estimating HLA/disease association with application to a study of aplastic anemia. Biometrics 1991; 47:53–61.

(12) Falk CT, Rubinstein P. Haplotype relative risks: an easy reliable way to construct a proper control sample for disk calculations. Ann Hum Genet 1987;51:227–33.

(13) Ott J. Statistical properties of the haplotype relative risk. Genetic Epidemiol 1989;6:127–30.

(14) Terwilliger JD, Ott J. A haplotype based haplotype relative risk approach to detecting allelic associations. Hum Hered 1992;42:337–46.

(15) Knapp M, Seuchter SA, Baur MP. The haplotype-relative-risk (HRR) method for analysis of association in nuclear families. Am J Hum Genet 1993;52:1085–93.

(16) Schaid DJ, Sommer SS. Genotype relative risks: methods for design and analysis of candidate-gene association studies. Am J Hum Genet 1993;53: 1114–26.

(17) Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506–16.

(18) Tiret L, Nicaud V, Ehnholm C, Havekes L, Menzel HJ, Ducimetiere P, et al. Inference of the strength of genotype-disease association from studies comparing offspring with and without parental history of disease. Ann Hum Genet 1993;57:141–9.

(19) Schaid DJ. General score tests for associations of genetic markers with disease using cases and their parents. Genet Epidemiol 1996;13:423–49.

(20) Lubin JH, Gail MH. Biased selection of controls for case–control analyses of cohort studies. Biometrics 1984;40:63–75.

(21) Greenland S, Thomas DC. On the need for the rare disease assumption in case–control studies. Am J Epidemiol 1982;116:547–53.

(22) Anton-Culver H, Kurosaki T, Taylor TH, Gildea M, Brunner D, Bringman D. Validation of family history of breast cancer and identification of the BRCA1 and other syndromes using a population-based cancer registry. Genet Epidemiol 1996;13:193–205.

(23) Curtis D, Sham PC. A note on the application of the transmission disequilibrium test when a parent is missing. Am J Hum Genet 1995;56:811–2.

(24) Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 1998;62:450–8.

(25) Whittemore AS, Halpern J. Multi-stage sampling in genetic epidemiology. Stat Med 1997;16:153–67.

(26) Siegmund KD, Whittemore AS, Thomas DC. Multistage sampling for disease family registries. Monogr Natl Cancer Inst 1999;26:43–8.

(27) Gail MH, Pee D, Carroll R. Kin-cohort designs for gene characterization. Monogr Natl Cancer Inst 1999;26:55–60.

(28) Zhao LP, Hsu L, Davidov O, Potter J, Elston R, Prentice RL. Population-based family study designs: an interdisciplinary research framework for genetic epidemiology. Genet Epidemiol 1997;14:365–88.

(29) Zhao LP, Aragaki C, Hsu L, Potter J, Elston R, Malone KE, et al. Integrated designs for gene discovery and characterization. Monogr Natl Cancer Inst 1999;26:71–80.

Note

Supported by Public Health Service grants CA58860 (National Cancer Institute) and 5P30 ES07048-03 (National Institute of Environmental Health Services) (W. J. Gauderman and D. C. Thomas), and CA73270 (National Cancer Institute) and RR03655 (National Center for Research Resources) (J. S. Witte), National Institutes of Health, Department of Health and Human Services.

Cohort Studies for Characterizing Measured Genes

Bryan Langholz, Nathaniel Rothman, Sholom Wacholder, Duncan C. Thomas

We describe the advantages of using established cohort studies that have collected blood samples to investigate the role of genes in the etiology of cancer. These studies include the cost-efficiency and reliability of nested case–control substudies from the cohort for exploration of gene–disease associations and gene–environment interactions as well as gene penetrance. Also, the cohort may serve as a well-defined "mini-population" from which to study population stratification and molecular markers of ethnicity. We conclude that cohort studies can play a significant role in assessing the role of genetic markers for common tumors or multiple cancer sites. [Monogr Natl Cancer Inst 1999;26:39–42]

Many open questions exist about the importance of genetics in causing cancer and about the degree to which genetic variation can explain the variation in cancer risk across different populations. Estimates of familial risk and penetrance for putative major genes can be obtained from family studies via segregation analysis *(1,2)* or, once localized, via joint segregation and linkage analysis *(3)*. Once the gene has been cloned, it is of interest to assess the role of the gene in cancer etiology in the general population. In some circumstances, family studies can be used to estimate penetrance *(4)*. But, to assess relative and attributable risk in the general population, population case–control studies have been the main approach used over the past decade to evaluate the effects of common genetic polymorphisms on cancer risk *(5)*. Case–control studies are particularly appealing when large numbers of cases need to be rapidly accrued or when time-consuming and resource-intensive exposure assessment is required. In studies of genetic factors, these advantages of the case–control design are not reduced by potential information biases, which may occur more commonly for certain exposures (i.e., dietary) assessed in case–control studies than in prospective cohort studies.

Cohort studies may also be used to evaluate genetic effects and may have some important advantages over case–control studies. Even if a case–control study might be more appropriate for answering a question about multiple exposures obtained from questionnaires about a single outcome quickly and efficiently, the cohort study allows one to study multiple outcomes with only laboratory costs hindering collection of information on many loci. Furthermore, although case–control designs will necessarily be the method of choice for studying very rare diseases, cohort studies currently under way are of sufficient size to yield enough cases to efficiently test genetic markers of common tumors, of multiple cancer sites, and, potentially, of important noncancer outcomes that can be assessed accurately *(6–8)*.

Table 1 lists many of the world's major prospective, relatively general cohort studies that have collected comprehensive food-frequency questionnaires and have or are collecting blood samples on at least 10 000 adults. Each of these cohort studies also has, to a variable extent, questionnaire data on other lifestyle and demographic characteristics, including race, ethnicity, and family history. At the completion of ongoing collections, blood samples will be stored on about one million individuals in these studies. Although samples from every study listed may not necessarily be available for investigations of genetic factors, this table does demonstrate the feasibility of collecting and storing biologic samples from subjects in large epidemiologic cohort studies.

Even within a cohort, one would capitalize on the efficiency of the case–control approach to investigate gene–disease associations. The cohort would serve as a study base for case–control studies in which only the blood from case and sampled control subjects would be assayed for genotype. In this paper, we discuss how this approach can be used to advantage in the study of gene–disease associations as well as how cohort studies may be used to address other hypotheses about the genetic etiology of cancer.

Study Methods

The starting point for studies of genetic factors within a cohort study is, of course, a relatively large number of individuals for whom blood or other biologic material has been collected and stored. This cohort is followed over time for occurrence of disease. Once enough cases of disease have occurred, case–control studies of genetic factors could be performed, with controls appropriately selected from the cohort members. These methods include nested case–control [e.g., *(9–11)*] and case–cohort [e.g., *(12,13)*] designs. (To simplify the terminology, we will only refer to nested case–control studies, but our discussion applies to case–cohort designs as well.) The biologic material from the subjects in this case–control study would be analyzed to determine genotype, and association studies of the genetic factor would proceed. The feasibility of this approach is based on the ability to store and retrieve biologic samples for a large number of people to reach the required number of cases but to genotype a much smaller number of subjects. We consider some potential advantages of this approach, in particular, in comparison to population-based, case–control studies.

Advantages of Nested Case–Control Studies From an Established Cohort

Nested Case–Control Studies Can Be Done Quickly

Enrolling subjects to participate in the cohort study; building the baseline database; collecting, cataloging, and storing biologic materials; and developing follow-up procedures to update disease event and time-varying information takes considerable

Affiliations of authors: B. Langholz, D. C. Thomas, Department of Preventive Medicine, University of Southern California, Los Angeles; N. Rothman, S. Wacholder, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

Correspondence to: Bryan Langholz, Ph.D., Department of Preventive Medicine, School of Medicine, University of Southern California, 1540 Alcazar St. CHP-220, Los Angeles, CA 90033 (e-mail: langholz@hsc.usc.edu).

See "Note" following "References."

© Oxford University Press

Table 1. Selected cohort studies with blood sample collections and food-frequency questionnaire data*

Study†	Year blood collection began	No. of subjects with blood samples
NYU Women's Health Study, United States	1985	14 000
Northern Sweden Health and Diseases Study, Sweden	1985	~43 000
ATBC, Finland	1985	29 000
Health Professional's Follow-up Study, United States	1986	18 000
ORDET, Italy	1987	11 000
ARIC, United States	1987	16 000
Nurses' Health Study, United States	1989	~33 000
Washington County, Maryland, United States	1989	33 000
Melbourne Collaborative Cohort Study, Australia	1990	42 000
JPHC, Japan	1990	49 000
Nurses' Health Study II, United States	1991	30 000
Women's Health Study, United States	1992	27 000
Women's Health Initiative, United States	1993	164 000
EPIC, Europe	1993	350 000
PLCO Study, United States‡	1994	~65 000
Shanghai Women's Health Study, China§	1997	~55 000
CPS-II Lifelink, United States‖	1998	~40 000

*Modified from Willett W. Nutritional epidemiology. 2nd ed. New York (NY): Oxford University Press; 1998. p. 486–7.

†NYU = New York University; ATBC = Alpha-Tocopherol Beta-Carotene Cancer Prevention Study Group; ORDET = Hormones and Diet in the Etiology of Breast Tumors Study; ARIC = Atherosclerosis Risk in Communities Study; JPHC = Japan Public Center-based prospective study on cancer and cardiovascular diseases; EPIC = European Prospective Investigation into Cancer and Nutrition; PLCO = Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; and CPS-II Lifelink = American Cancer Society Cancer Prevention Study-II Lifelink Cohort.

‡Hayes R: personal communication (planned number, still enrolling).

§Zheng W, Chow WH: personal communication (planned number, still enrolling).

‖Thun M: personal communication (planned number, still enrolling).

time, money, and effort. But once the cohort study resource is established and a sufficient number of cases have occurred, a study of genetic factors can proceed much more quickly and efficiently than a population-based study. At this stage, the registry has already ascertained cases, and control selection from within the cohort is a statistical and administrative activity. The main effort at this point is the laboratory work needed to determine genotypes. In contrast, a population-based, case–control study requires ascertainment and enrollment of diseased subjects, as well as blood collection, before reaching the laboratory analysis stage. Thus, once the cohort study is established and matured, the additional cost of a nested case–control study is relatively small, and, perhaps more important, the time needed to do the studies can be much shorter. Given this advantage, nested case–control studies provide a timely and reliable way to verify gene–disease associations found in other studies.

Multiple Studies From the Cohort Are Comparable

The cohort resource permits case–control studies of different disease outcomes, all from the same study base. The ability to efficiently study multiple tumors might compensate for their poorer efficiency for a single hypothesis. In addition, the outcomes can be reliably compared, if case ascertainment is comparable. For this purpose, a case–cohort study design should be considered because controls can be shared among the various outcomes *(14)*.

Nested Case–Control Studies Are Relatively Free of Selection Bias

It can be difficult to get controls in population-based, case–control designs that adequately represent the base population, and seldom is there any way to assess their representativeness. In a cohort study, the cohort is the base population, so random sampling from it is straightforward. Even if one cannot obtain genotypes on all of them, one can at least characterize the losses in terms of baseline characteristics.

Gene by Environment Studies

Gene–environment interactions can be explored in cohort studies with the use of relatively unbiased exposure data assessed biologically or by questionnaire that can potentially be collected at more than one point in time. Under these circumstances, the nested case–control study would avoid recall or other information biases that can occur when "environment" data are retrospectively obtained by interview, as is usually the situation with population-based, case–control studies. We note that misclassification of exposure status can seriously bias the assessment of gene–environment interactions and substantially reduce power *(15)*. If nested case–control study subjects need to be contacted to obtain additional exposure information, contact information will be available to the study investigators. Also, participation rates will be high because cohort members have already agreed to participate in the main cohort study. Of course, interview data collected retrospectively will be subject to the same information biases as a population-based study and, to the extent that cases have died or refuse to participate, selection biases. Also, there can be added selection bias from loss to follow-up if censoring is informative.

Penetrance Estimation

Because the study population is enumerated, the penetrance of the genes can be reliably estimated from the nested case–control study, accounting for risk factors measured on the case–control subjects *(16)*. In this regard, such studies could be definitive about penetrance for populations that are reasonably represented, in terms of unmeasured risk factors, by the cohort. Absolute risk can be estimated from population based, case–control studies but will be subject to much more error because of unidentified cases and because of inaccuracies in the estimation of numbers in the underlying population *(17)*.

Penetrance estimation from family-based studies requires strong assumptions to extrapolate the prevalence of the gene from the included families to the general population. Although methods can, in principle, correct for ascertainment *(18)*, these methods may still lead to upwardly biased estimates if the penetrance is not the same in all families *(19)*.

Cost-Efficient Sampling Schemes

As described above, nested case–control studies from a cohort are a cost-efficient study design in the same way that case–control studies from a population are cost-efficient. To understand the relative contribution of individuals with respect to information about gene characterization from the cohort, it is useful to think of the cross-classification of four states on the basis of presence or absence of the gene and disease status *(20)*.

The relative size of these cells is proportional to the amount of information contributed by individuals in these cells. Thus, if the disease is rare, the information contributed by each diseased subject is far greater than from a nondiseased subject, so that virtually all of the information about the gene–disease association may be captured with the use of information from a small fraction of the nondiseased subjects. This, of course, is the basis for case–control study designs. Disease status information that is available on all cohort members is used in sampling from the cohort, so that costly information (genotype) needs only to be collected on a fraction of the entire cohort to capture nearly all of the information in the cohort about gene–disease association. If the gene is rare, then the individuals with gene-positive cells will carry relatively large amounts of information. Although, of course, genotype status is not known for the cohort subjects, some factor that is correlated with the presence of the gene, such as family history, or a positive value from an inexpensive lab assay, may be available. Just as disease status information may be exploited to increase the information per sampled subject, gene-related information can be used to increase the information from each sampled subject and to decrease the sample size, and thus the cost, of the study. Counter-matching and other two-stage designs that exploit gene-related information on cohort members may reduce the number of subjects who must be genotyped, relative to standard case–control studies *(21–25)*.

Other Uses for Cohort Studies

We have thus far focused on case–control substudies within the cohort. But an established cohort study can be useful in other study designs and also can be used to address issues other than gene–disease associations.

Opportunities to Expand the Usefulness of the Cohort by Adding Information From Family Members

In addition to exploiting the information available on the cohort members themselves, it may be feasible in some circumstances to use them as probands for family-based, cohort or case–control studies [e.g., *(26,27)*]. For example, one might use the cases from the cohort and a random or matched sample of unaffected controls as probands in the kin-cohort design *(19)*. In this approach, only the cohort members would require genotyping. The only additional data to be obtained would be the subjects' family histories in first-degree relatives, which might be already available from the probands' baseline questionnaires or might be obtained by follow-up questionnaires to the probands, without the need to formally enroll their relatives. These designs could also be extended to include more distant family members or to obtain genotypes from selected relatives. Multistage sampling designs might be useful in this context *(24)*.

Conceptually, these family-based designs are no different from those discussed elsewhere in this workshop—sampling probands from the general population. However, given the need to involve family members, the availability of already collected family history information on a baseline questionnaire could considerably simplify their implementation, particularly for designs restricted to multiple-case families. For example, in the Department of Preventive Medicine, University of Southern California, Los Angeles, a sibling case–control study of breast and ovarian cancers nested is currently being conducted within a multiethnic cohort study in Los Angeles and Hawaii. All sibships with at least two cases (in either the sibship or their parents) and at least one control are being enrolled, this information being readily available from the baseline family-history questionnaire. In addition, being able to characterize the representativeness of the case series vis-à-vis the larger cohort would help address concerns about possible selection biases in such designs.

Opportunities to Explore Population Stratification Bias

Some investigators *(26,28,29)* have voiced concerns about the potential for bias in case–control studies with unrelated controls because of population stratification. Although diabetes in the Pima Indians provides one notable example of such bias *(30)*, the extent of the problem in general is still unresolved. Cohort studies that use unrelated individuals are, in principle, subject to exactly the same concerns. However, baseline information on ethnicity and other risk factors that are sometimes available on the entire cohort provides opportunities to examine the potential severity of this problem in ways that may not be feasible when sampling from the general population without a well-defined sampling frame. In other words, by treating the cohort as a "mini-population," one could contrast the results from alternative case–control designs (e.g., using unrelated and family member controls) in terms of their ability to correctly estimate the parameter that would have been obtained by using the full cohort and to examine the role of the available ethnic or other baseline data to account for methodologic differences. The cohort results could then be used to infer the potential effect of population stratification in other studies on the basis of the ethnicity and other risk factors in the study population and the type of study design used.

Control of Ethnic Stratification With the Use of Molecular Markers

A relatively unexplored approach to the ethnic stratification problem entails the use of polymorphic markers to infer ethnicity. Shriver et al. *(31)* have proposed an approach to inferring the ethnic affiliation of individuals with the use of a panel of markers whose allele frequencies vary substantially between ethnic groups. By using a similar panel with a maximum likelihood approach to discrimination, Shriver et al. demonstrated reasonable separation in the likelihood scores between European Caucasians and African-Americans but less discrimination for Hispanics. To date, we are not aware of any applications of this approach to ethnic stratification in epidemiologic studies of candidate gene associations. However, Pritchard and Rosenberg *(32)* discuss the potential of this approach for detecting population stratification and provide guidelines as to the number of unlinked markers that would be needed. Although this idea could be implemented in a population based, case–control study, the approach might be most easily implemented in a cohort study by first picking a few ethnically matched controls per case and, as part of the genetic typing, ascertain the alleles for markers of "ethnic origin." A further analysis of cases with controls of the most closely matching ethnicity would provide an indication of the extent that population stratification might explain an observed gene–disease association.

Discussion

Genetic studies within a cohort study are feasible when blood or other genetically analyzable material is available on study members and enough cases of disease are (or will be) available to ensure statistical precision. We have suggested that case–

control studies within the cohort would be the design of choice for gene-association studies because blood from only a small fraction of the cohort members would need to be genotyped. We have outlined a number of benefits of this approach, including low marginal cost and high reliability of such studies. Also, they could be used to investigate familial clustering of disease and population stratification issues.

Beside the restriction to relatively common diseases and genes, some other potential limitations exist to using cohort studies for genetic studies. Cohorts are often not representative of the "general population." This lack of representation may be because the cohort is formed precisely because it is a "high-exposure" group, such as in many occupational cohorts. Or subjects are enrolled from a convenient administrative entity, such as unions, or religious or professional organizations. Furthermore, those who choose to participate and, in particular, agree to provide a blood sample can ultimately be a rather small and selected proportion of those approached, thus, perhaps, limiting the generalizability of the estimated associations and risks. Cohort study reliability depends on many data-quality issues, such as the quality of questionnaire and other information gathered in the field, completeness of follow-up, and outcome ascertainment. In particular, if loss to follow-up is related to presence of the gene, bias can arise. If loss to follow-up is significant and differential across a gene-related factor (e.g., race or ethnicity), some protection from such bias would be obtained by stratification of the analysis by those factors because the follow-up will be less differential within these strata.

Mounting and maintaining a cohort study is a major endeavor. But, once established, the cohort study can play an important role by quickly and definitively verifying gene–disease associations and, potentially, gene–environment interactions as well as establishing the penetrance of the gene.

References

(1) Schwartz AG, King MC, Satariano WA, Swanson GM. Risk of breast cancer to relatives of young breast cancer patients. J Natl Cancer Inst 1985;75:665–8.

(2) Elston RC, Rao DC. Statistical modeling and analysis in human genetics. Annu Rev Biophys Bioeng 1978;7:253–86.

(3) Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Am J Hum Genet 1995;56:265–71.

(4) Gail MH, Pee D, Benichou J, Carroll R. Designing studies to estimate the penetrance of an identified autosomal dominant mutation: cohort, case–control, and genotype-proband designs. Genet Epidemiol 1999;16:15–39.

(5) Caporaso N, Rothman N, Wacholder S. Case–control studies of common alleles and environmental factors. Monogr Natl Cancer Inst 1999;26:25–30.

(6) Rothman N, Stewart W, Schulte PA. Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. Cancer Epidemiol Biomarkers Prev 1995;4:301–11.

(7) Munoz A, Grange SJ. Methodological issues for biomarkers and intermediate outcomes in cohort studies. Epidemiol Rev 1998;20:29–42.

(8) Khoury MJ, Beaty TH, Cohen BH. Fundamentals of genetic epidemiology. New York (NY): Oxford University Press; 1993.

(9) Mantel N. Synthetic retrospective studies and related topics. Biometrics 1973;29:479–86.

(10) Lubin J, Gail M. Biased selection of controls for case–control analysis of cohort studies. Biometrics 1984;40:63–75.

(11) Langholz B, Goldstein L. Risk set sampling in epidemiologic cohort studies. Stat Sci 1996;11:35–53.

(12) Kupper L, McMichael A, Spirtas R. A hybrid epidemiologic study design useful in estimating relative risk. J Am Stat Soc 1975;70:524–8.

(13) Prentice RL. A case–cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 1986;73:1–11.

(14) Wacholder S. Practical considerations in choosing between the case–cohort and nested case–control designs. Epidemiology 1991;2:155–8.

(15) Rothman N, Garcia-Closas M, Stewart WS, Lubin J. The impact of misclassification in case–control studies of gene–environment interactions. In: Vineis P, Malats N, Lang M, d'Errico A, Carporaso N, Cuzick J, et al., editors. Metabolic polymorphisms and susceptibility to cancer. Lyon (France): International Agency for Research on Cancer (IARC); 1999.

(16) Langholz B, Borgan O. Estimation of absolute risk from nested case–control data. Biometrics 1997;53:767–74.

(17) Benichou J, Gail M. Methods of inference for estimates of absolute risk derived from population-based case–control studies. Biometrics 1990;51:182–94.

(18) Risch N. Segregation analysis incorporating linkage markers. I. Single-locus models with an application to type I diabetes. Am J Hum Genet 1984;36:363–86.

(19) Wacholder S, Hartge P, Struewing JP, Pee D, McAdams M, Brody L, et al. The kin-cohort study for estimating penetrance. Am J Epidemiol 1998;148:623–30.

(20) Wacholder S. Design issues in case–control studies. Stat Meth Med Res 1995;4:293–309.

(21) Langholz B, Borgan O. Counter-matching: a stratified nested case–control sampling method. Biometrika 1995;82:69–79.

(22) Langholz B, Clayton D. Sampling strategies in nested case–control studies. Environ Health Perspect 1994;102:47–51.

(23) Goldstein AM, Andrieu N. Detection of interaction involving identified genes: available study designs. Monogr Natl Cancer Inst 1999;26:49–54.

(24) Siegmund KD, Whittemore AS, Thomas DC. Multistage sampling for disease family registries. Monogr Natl Cancer Inst 1999;26:43–8.

(25) Borgan O, Langholz B, Samuelsen SO, Goldstein L, Pogoda J. Exposure stratified case–cohort studies. Lifetime Data Anal. In press 1999.

(26) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Monogr Natl Cancer Inst 1999;26:31–7.

(27) Gail MH, Pee D, Carroll R. Kin-cohort designs for gene characterization. Monogr Natl Cancer Inst 1999;26:55–60.

(28) Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994;265:2037–48.

(29) Witte JS, Gauderman WJ, Elston RC, Thomas DC. Bias and efficiency in case–control studies of candidate genes and gene–environment interactions: basic family designs. Am J Epidemiol 1999;149:693–705.

(30) Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 1988;43:520–6.

(31) Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R. Ethnic-affiliation estimation by use of population-specific DNA markers. Am J Hum Genet 1997;60:957–64.

(32) Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 1999;65:220–8.

Note

Supported by Public Health Service grants CA42949 and CA52862 (B. Langholz and D. C. Thomas) from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

Multistage Sampling for Disease Family Registries

Kimberly D. Siegmund, Alice S. Whittemore, Duncan C. Thomas

Background: **The objectives of a family-based disease registry range from characterizing measured genetic factors and gene–environment interaction effects to detecting novel susceptibility genes. Gathering complete information on exposure and disease status in all family members for a sample of affected subjects (probands) to address these diverse objectives would be prohibitively expensive.** ***Methods:*** **Multistage sampling can be used to design an efficient family-based disease registry. At each stage, the probands are classified on the basis of previously collected data, and a subsample is selected for more detailed observation. The design can be optimized to minimize the variance of any of the model parameter estimates, subject to a constraint on the total sample size.** ***Results:*** **We describe the basic statistical theory and its application to a four-stage sampling scheme proposed for the Cooperative Family Registry for Epidemiologic Studies of Colorectal Cancer at the University of Southern California. [Monogr Natl Cancer Inst 1999;26:43–8]**

Several papers in this monograph describe a variety of basic designs for characterizing genes and gene–environment interactions. These designs include case–control studies that use unrelated controls from the general population *(1)*, family-based case–control designs *(2)*, family-based cohort designs *(3)*, and variants of these designs for testing interaction effects *(4)*. However, all of these basic designs for rare diseases may still be relatively inefficient for studying genes with rare mutations like BRCA1 because the yield of carriers in a random sample of case patients and control subjects will be very low, even if the genetic relative risk is high. To overcome this difficulty, a variety of approaches has been proposed, based on the restriction of samples to case patients and control subjects with a positive family history, who will tend to have a higher carrier frequency than an unselected series of individuals. For example, D. C. Thomas (unpublished data) has examined the relative efficiency of various family-based case–control designs incorporating such restrictions [*see also (2)*]. Nevertheless, with increasing degree of restriction, the relevance of the findings from such a study to the general population becomes more questionable, particularly if there may be other unmeasured environmental or genetic risk factors that modify the effect of the genes under study. As an alternative, we discuss a class of designs that entails random sampling of case patients from all strata of family history in such a way as to allow population-based inference while at the same time maximizing the statistical efficiency of the design. (Here, an optimal design minimizes the variance for any of the parameter estimates, such as the probability of disease given genotype [penetrance] or the frequency of the variant allele, under a fixed total sample size.) To address these goals, we explore various multistage designs, involving stratified random sampling at each stage with the use of family history data that are collected at each successive stage.

The basic idea of multistage sampling was introduced into epidemiology by White *(5)* in the context of a case–control design for studying the association between a rare disease and a rare exposure. Since then, there has been an abundance of related work by others *(6–11)*. An elementary two-stage design is the following. In the first stage, one selects a random sample of case patients and control subjects and assesses exposure on all of them; in the second stage, one conducts a subsample from the 2 × 2 table of exposed and unexposed case patients and control subjects and collects additional covariate information only on these sampled individuals. Alternatively, one might obtain a surrogate for exposure on all participants in the first stage as the basis for stratification and then obtain more precise exposure data only on participants in the second stage. The latter is particularly relevant in the genetic context in which family history is a natural surrogate to consider for the genotypes that might be measured in later stages but would be too expensive to obtain for all participants.

Whittemore and Halpern *(12)* have discussed the application of multistage sampling to the field of genetic epidemiology. As one example, they described a sampling design for a study of prostate cancer. In stage 1, a case–control study using population control subjects was conducted to evaluate the association of diet and other lifestyle characteristics with prostate cancer. A brief family history of disease in fathers and brothers was obtained on all participants in this stage. In stage 2, subsets of the case patients and control subjects were selected conditional on their family history. From these selected participants, additional data on disease status in family members and their ages at onset of prostate cancer or censoring together with medical verification of the reported cancers were collected for segregation analysis. Families that contained three or more medically verified cases with prostate caneer were pursued in stage 3 for DNA samples of blood or tissue specimens, or a combination of both, for linkage analysis. With the use of simulation, Whittemore and Halpern determined the sampling fractions for the second stage that would minimize the variance of the penetrance and allele frequency estimates from the segregation analysis and maximize the yield of carriers for subsequent linkage analysis. They considered an initial sample of 1500 case patients and 1500 control subjects in stage 1 and a subsample of 570 participants in stage 2. In stage 1, 209 case patients and 92 control subjects had a positive family history of disease. Their results indicated that, regardless of which parameter they were estimating, the optimal stage 2 design entailed selecting all available case patients with a positive family history (n = 209) and subsamples of case patients and control subjects from the remaining categories of family history and disease status. The optimal fraction in those

Affiliations of authors: K. D. Siegmund, D. C. Thomas, Department of Preventive Medicine, University of Southern California, Los Angeles; A. S. Whittemore, Department of Health Research and Policy, Stanford University School of Medicine, CA.

Correspondence to: Kimberly D. Siegmund, Ph.D., Department of Preventive Medicine, University of Southern California, 1540 Alcazar St., Suite 220, Los Angeles, CA 90089-9011 (e-mail: kims@rcf.usc.edu).

See "Note" following "References."

© Oxford University Press

remaining categories varied, depending on which parameter was of greatest interest. The greatest difference in sampling fractions was observed for estimating the frequency of the variant allele: A large fraction of family history-positive control subjects (64%) were desired as well as much smaller fractions of family history-negative case patients (20%) and family history-negative control subjects (3%). For optimizing the estimate of the hazard rate in noncarriers, a smaller fraction of family history-positive control subjects (34%) and family history-negative cases (12%) was required, as was a greater fraction of family history-negative control subjects (13%). The optimal sampling fractions to estimate the hazard rate in carriers lay between the extremes given by the other two parameters. For linkage analysis, the optimal design involved maximizing the yield of carriers in the sample. This design entailed subsampling the pedigrees having the highest probability of segregating the gene: all family history-positive case patients and control subjects and 21% of the family history-negative case patients. Because the sample sizes vary across the different strata in the first stage, the largest group in absolute numbers is sometimes the family history-negative case patients.

In the following section, we summarize the basic statistical theory. We then discuss the application of these principles to the design of the National Cancer Institute Cooperative Family Registry for Colorectal Cancer Research at the University of Southern California. This registry has diverse objectives, including characterization of genes that may interact with environmental exposures as well as to provide a resource for locating new genes that influence disease risk. Inevitably, no single design can be optimal for all of these objectives, so we discuss the compromises involved and the design we finally recommended. This design involves three key differences from that discussed by Whittemore and Halpern *(12)*. First, the sample will be based only on families of case patients, with no families of control subjects. Second, we will have available genotype data on the case patients for the purpose of estimating penetrance and allele frequency, rather than relying purely on segregation analysis methods. Third, we will also have genotypes and risk-factor information available on selected family members for the purpose of studying gene–environment interaction effects.

Statistical Methods

The methods for characterizing disease genes are likelihood based (e.g., conditional logistic regression and segregation analysis). If resources permitted sampling all probands and their family members in a single stage, we would have a sample of independent families. Consistent parameter estimates could then be obtained by solving the score equation (setting the derivative of the log-likelihood to zero), and parameter variances could be estimated, using the Fisher information in the usual way. Instead, we compute the variance of parameter estimates for various two-stage designs in which only a stratified subsample of families is selected. To compare the efficiencies of two designs having equal sample sizes, we compute the asymptotic relative efficiency, the ratio of the inverse variance estimates.

Appropriate statistical methods for data analysis use the information collected in stage 1 to account for differential sampling across strata. Under maximum likelihood, the score equation is a function of both the data observed at the second stage and the distribution of the stratification variable. This latter distribution may be unknown or difficult to calculate. Although the distribution is not of real interest, its misspecification may lead to bias in our parameter estimates. An alternative method, which is not susceptible to this bias, is the Horvitz–Thompson approach *(13–16)*. The Horvitz–Thompson estimating equation is the score equation using only the data observed in the second stage but weighted by the inverses of the sampling fractions. To be well defined, this method requires that all strata be represented in stage 2. The variance estimator can be written as the sum of two parts: the variance of the maximum likelihood estimator based on the complete data and a penalty term for the loss of precision as a result of sampling. This latter term depends on the number of observations in each stratum, the sampling fractions, and the variability of the score function within the strata. The equations for the score and variance are given in Appendix A. To optimize the two-stage design, we minimize the asymptotic variance of the Horvitz–Thompson estimates.

The Horvitz–Thompson approach can be used for estimating population-based parameters from a two-stage sample taken from traditional genetic and epidemiologic study designs. In stage 1 of a family study, we sample affected patients (probands) and classify them according to history of disease in their first-degree relatives (e.g., 0, 1, 2+ members affected). In stage 2, we sample case probands at random from within each stratum and collect extended pedigree data out to first cousins. The penetrance and allele frequency for an unobserved gene in such samples may be estimated with the use of classical segregation analysis models that account for the disease status of the sampled proband *(12)*. The likelihood contribution for a single family is computed as the ratio of the probability of disease in the family over the probability that the proband is affected. The probability of disease in the family is obtained by summing the joint probability of disease status (d) and genotypes (g) over all possible genotypic combinations. When genotypes for some individuals are observed, we can include the probability of observing those genotypes in our model for a modified segregation analysis. Let the subscripts o and u denote the observed and unobserved genotypes, respectively, then the numerator of the likelihood can be written as

$$L(\theta;d,g_o) = \sum_{g_u} \Pr(d,g_o|g_u)\Pr(g_u)$$

where θ denotes the vector of model parameters (carrier and noncarrier penetrances and allele frequency). In theory, this model could be extended to include data on environmental exposure if such data are collected on all family members. However, because it will only be feasible to collect risk factor data on a subset of living family members and comparisons of risk factors would only be meaningful between individuals in the same generation, we resort to conditional logistic regression for family-matched, case–control sets.

Family-matched, case–control sets are used to estimate the relative risk of disease by environmental exposure and genotype and the ratio of relative risks for gene-by-environment interaction effects. The dominance effect of genes can be modeled as recessive, dominant, or multiplicative on the relative risk scale (*see* Appendix B). With the use of family-matched risk sets, the distribution of families in the stage 1 strata does not contribute to the conditional likelihood. Furthermore, the stage 2 sampling fractions cancel from the likelihood for case–control sets matched on the history of disease in the family. Therefore, a control subject that is a sibling to the case patient always shares

the same family history of disease as the case patient, and the analysis is straightforward using ordinary conditional logistic regression. However, a control subject that is a cousin to the case patient will not necessarily share the same family history (e.g., affected mother), and it is again necessary to account for the sampling fractions in the likelihood for the resulting estimates to be valid.

Association or linkage analysis can be used to locate new genes as risk factors for disease. The same case–control methods described above can be used for the purpose of detecting associations of new candidate loci or loci in linkage disequilibrium with unidentified causal genes. Alternatively, linkage analysis can be used, including both model-based methods that describe the transmission of genes in families or model-free methods based on the sharing of ancestral marker alleles. The optimal design for model-based linkage analysis is to maximize the number of carriers in the sample. To accomplish this, we subsample the pedigrees that have a high probability of segregating the gene. We calculate this probability for each stratum and then select families with the largest probability of segregating the disease genotype followed by selecting those with the second largest and so on until we reach our desired sample size. Such pedigrees could later be extended to include more distant relatives, following the sequential sampling method of Cannings and Thompson *(17)*.

Application

The goals of the Cooperative Family Registries for Breast and Colorectal Cancer Research include developing a resource of family data for use in both gene characterization and gene hunting. For gene characterization, distinctions are made between major genes and metabolic genes. We refer to a single gene that has a large relative risk of disease because of a rare allele as a major gene. Metabolic genes, conversely, denote genes that have a small relative risk and the disease-predisposing allele is common in the population (e.g., frequency 20% or more). For the genes with rare alleles, attention focuses on estimating allele frequency and absolute penetrance in carriers and noncarriers. For genes with common alleles, we focus instead on the relative penetrance. Detecting departures from the multiplicative effects of the relative risks for genotype and environmental exposure is of interest for both major and metabolic genotypes.

The primary design considerations include which families to select in a stratified sample of the ascertained case probands and from whom to collect blood samples and risk-factor questionnaires. For all designs, we assumed a genetic relative risk of 20 for a major gene and 2 for a metabolic gene as well as a population prevalence of disease of 10%. We expect to be able to ascertain 5000 probands from whom it will be feasible to sample 1000 families (20%) and to obtain blood samples from approximately 4500 individuals. We considered a fixed family structure consisting of the case proband, two siblings, mother, father, two first cousins, the cousins' mother and father (proband's aunt and uncle), and the connecting grandparents.

First, we computed the optimal sampling fractions for minimizing the variance of the segregation parameters for a variety of genetic models when the genotype of only the case proband is observed. In general, the results suggest only slightly oversampling families with 0 and 2+ affected relatives to minimize the variance of the allele frequency. For example, the optimal design for a rare recessive gene (allele frequency = 0.14) is to sample 20.4%, 18.5%, and 22.6% of case probands with 0, 1, and 2+ affected first-degree relatives, respectively. To minimize the variance of the penetrance in carriers and noncarriers, we should undersample case probands with 0 affected relatives and oversample those with 2+ affected relatives. For the same rare recessive gene, we would sample 17.3%, 18.3%, and 49.9% of case probands with 0, 1, and 2+ affected first-degree relatives, respectively, to efficiently estimate the penetrance in gene carriers. The averages of the fractions over the different parameters of interest are given in Table 1 (19%, 19%, 35% in stratum with 0, 1, 2+ affected relatives, respectively). The choice of parameterization of the penetrance as absolute or relative had no noticeable effect on the optimal design.

The asymptotic relative efficiency results comparing designs with different genotyped family members for a rare major gene and common metabolic gene are given in Tables 1 and 2, respectively. The greatest gain in efficiency (per individual) is seen when the proband is genotyped. For a rare major gene, the genotyping of additional family members increases efficiency the most in estimating the hazard rate in carriers and the allele frequency. Efficiency is also improved with additional genotyping for estimating the relative risk of a metabolic gene (Table 2). The largest efficiency gain for the relative risk is observed under the multiplicative model. Compared with genotyping the proband only, the relative efficiency of additionally genotyping both siblings is 3.5 and of additionally genotyping both parents is 4.5.

Asymptotic relative efficiency results for gene–environment interaction effects in case–control sib pairs suggest the greatest efficiency is obtained from sampling pairs that have a positive

Table 1. Asymptotic relative efficiency of the Horvitz–Thompson estimates for designs in which different genotypes are observed (genetic relative risk = 20)*

Observed genotype(s)	Average sampling fractions, % ($\bar{f}_0, \bar{f}_1, \bar{f}_{2+}$)†	ARE‡		
		λ_0	λ_1	q
Recessive gene (q = 0.14)				
None	15, 18, 70	0.10	0.05	0.01
Proband	19, 19, 35	1.00	1.00	1.00
Proband + 2 siblings	18, 21, 29	1.08	2.51	2.07
Proband + 2 parents	18, 21, 30	1.08	2.53	2.44
Multiplicative gene (q = 0.02)				
None	16, 18, 64	0.04	0.03	0.01
Proband	19, 18, 35	1.00	1.00	1.00
Proband + 2 sibs	19, 20, 33	1.09	1.66	1.69
Proband + 2 parents	18, 21, 32	1.10	1.85	1.97
Dominant gene (q = 0.01)				
None	14, 21, 49	0.27	0.22	0.13
Proband	17, 22, 30	1.00	1.00	1.00
Proband + 2 sibs	17, 23, 31	1.04	1.43	1.26
Proband + 2 parents	17, 22, 32	1.04	1.31	1.27

*Population prevalence of disease = 10%. The sample contains 5000 probands in stage 1, 1000 of which are sampled in stage 2.

†The average sampling fraction by stratum (0, 1, 2+ affected relatives); averages are computed over three designs; each design minimizes the variance of one of the three parameters estimated, subject to the constraint of 1000 stage 2 probands.

‡ARE = asymptotic relative efficiency; λ_0 = prevalence of disease in noncarriers; λ_1 = prevalence of disease in carriers; and q = frequency of variant allele.

Table 2. Asymptotic relative efficiency for the Horvitz–Thompson estimates for designs in which different genotypes are observed (genetic relative risk = 2)*

Observed genotype(s)	Sampling fractions, % (f_0, f_1, f_{2+})†	Log_e (relative risk)
	Recessive gene (q = 0.44)	
None	12, 29, 67	0.003
Proband	18, 20, 37	1.00
Proband + 2 siblings	18, 22, 36	2.21
Proband + 2 parents	18, 22, 35	2.62
	Multiplicative gene (q = 0.19)	
None	14, 21, 85	0.001
Proband	17, 21, 55	1.00
Proband + 2 sibs	18, 21, 39	3.49
Proband + 2 parents	17, 23, 37	4.52
	Dominant gene (q = 0.10)	
None	11, 24, 94	0.021
Proband	16, 23, 47	1.00
Proband + 2 sibs	17, 22, 39	2.50
Proband + 2 parents	17, 24, 38	3.08

*Population prevalence of disease = 10%. The sample contains 5000 probands in stage 1, 1000 of which are sampled in stage 2.

†The optimal sampling fraction for the log relative risk parameter, by stratum (0, 1, 2+ affected relatives), subject to the constraint of 1000 stage 2 probands.

family history of disease (Table 3). The increase in efficiency is greater for the rare genotype than for the common one. Under both the multiplicative and dominant models, the efficiency gain from having an affected sibling is similar to that of having an affected parent. Under a recessive model, the increase in efficiency is greater when the affected relative is a sibling.

Finally, we find that, for all three dominance models, the probability of a family segregating a disease allele increases with the number of cases in the family (results not shown). Thus, the families with three or more cases will be the most informative for linkage analysis.

Table 3. Asymptotic relative efficiency for estimating gene-by-environment interaction effect in case-sib-control pair with an affected relative compared with an unaffected relative*

Relative risk ratio $G \times E$	Model	Allele frequency	ARE: Affected parent	ARE: Affected sibling
		Genetic relative risk = 20		
0.5	Recessive	0.14	1.75	2.74
	Multiplicative	0.02	1.99	2.00
	Dominant	0.01	3.90	3.91
1.0	Recessive	0.14	1.94	3.06
	Multiplicative	0.02	2.18	2.19
	Dominant	0.01	4.41	4.44
2.0	Recessive	0.14	2.09	3.31
	Multiplicative	0.02	2.37	2.37
	Dominant	0.01	4.82	4.85
		Genetic relative risk = 2		
0.5	Recessive	0.44	1.06	1.09
	Multiplicative	0.19	1.06	1.06
	Dominant	0.10	1.12	1.12
1.0	Recessive	0.44	1.14	1.20
	Multiplicative	0.19	1.15	1.15
	Dominant	0.10	1.26	1.27
2.0	Recessive	0.44	1.21	1.31
	Multiplicative	0.19	1.23	1.24
	Dominant	0.10	1.40	1.40

*Relative risk of environment = 2; ARE = asymptotic relative efficiency.

Our preferred "compromise" design, as recommended to the Colorectal Cancer Family Registry at the University of Southern California, is to sample the data in four stages.

In stage 1, we plan to enroll approximately 5000 probands with colorectal cancer from population-based cancer registries. In a short telephone interview, we will collect family history data on colorectal cancer in parents and siblings only and stratify the families on the number of affected parents and siblings.

In stage 2, a random sample of families will be drawn from each stratum. We propose selecting 16% of the probands with no family history of disease, 32% of the probands with one additional affected family member, 48% of those with two, 64% of those with three, etc. We anticipate sampling approximately 1000 families in stage 2. In an extended telephone interview with the proband, we will then inquire about cancer incidence in all first- and second-degree relatives and first cousins and request permission to contact all surviving affected family members, one unaffected sibling of each case, and two unaffected cousins of the proband. For probands without an unaffected sibling, we will seek permission to contact the parents. Blood samples will be drawn, and risk-factor and food-frequency questionnaires will be collected from the probands in this subsample.

In stage 3, blood samples as well as risk-factor and food-frequency questionnaires will be sought from relatives identified in stage 2 whom the proband has granted permission for us to contact. For probands without an affected sibling, parents will be sought for blood samples only. Alternatively, one could select relatives from a subsample of probands defined by cancer in the extended pedigree or by the genotype of the proband.

In a possible fourth stage, we will consider collecting extended pedigrees from families that are potentially informative for linkage analysis on the basis of their apparent segregation of disease that cannot be explained by already known genes, using the sequential sampling approach of Cannings and Thompson *(17)*.

The expected confidence intervals for the penetrance parameters and allele frequency for a major gene and a metabolic gene are given in Table 4. These expected confidence intervals are calculated, using the sampling fractions, the extended family history data, and the genotypes on the probands alone (approximately 1125). On the basis of the relative efficiency comparisons in Tables 1 and 2, we anticipate that, by including the genotypes of additional family members, the confidence intervals on the carrier penetrance, relative risk, and gene frequency could be substantially narrowed. For example, if we include the genotypes of two siblings for the common metabolic gene (approximately 3375 total genotypes), the sizes of the estimated confidence intervals for the genetic relative risk are reduced 35%–52%, depending on the genetic model. For a true genetic relative risk of 2, the estimated 95% confidence interval is 1.17–3.43 for a multiplicative susceptibility gene and approximately 1.43–2.79 for one acting in a dominant or a recessive fashion. In a separate case–control analysis of these same sib triplets, the estimated 95% confidence interval for a gene–environment relative risk ratio of 2 ranges from 0.7–5.8 for the

Table 4. Estimated 95% confidence intervals (CIs) on penetrance and allele frequency from the two-stage sample selecting 16%, 32%, and 48% of families with 0, 1, and 2+ affected relatives, respectively*

Model	Allele frequency	No. of families†	Noncarrier penetrance (asymptotic 95% CI)	Carrier penetrance (asymptotic 95% CI)	Allele frequency (asymptotic 95% CI)
			Genetic relative risk = 20		
Recessive	0.14	1155	0.089 (0.079–0.100)	0.661 (0.423–0.838)	0.14 (0.110–0.170)
Multiplicative	0.02	1145	0.091 (0.081–0.103)	0.668 (0.421–0.847)	0.02 (0.009–0.031)
Dominant	0.01	1203	0.089 (0.079–0.100)	0.660 (0.557–0.750)	0.01 (0.006–0.014)
Model	**Allele frequency**	**No. of families**	**Noncarrier penetrance (asymptotic 95% CI)**	**Relative risk (asymptotic 95% CI)**	**Allele frequency (asymptotic 95% CI)**
			Genetic relative risk = 2		
Recessive	0.44	1126	0.086 (0.073–0.101)	2.00 (1.23–3.25)	0.44 (0.37–0.51)
Multiplicative	0.19	1122	0.088 (0.069–0.110)	2.00 (0.73–5.46)	0.19 (0.11–0.27)
Dominant	0.10	1129	0.083 (0.073–0.102)	2.00 (1.18–3.38)	0.10 (0.05–0.15)

*n = 5000 stage 1 families; population prevalence of disease = 10%; genotype on proband only.

†Families consist of proband, two siblings, mother, father, two first cousins, the cousins' mother and father (proband's aunt and uncle), and connecting grandparents.

multiplicative gene model to 1.0–4.1 for the dominant and recessive model.

Discussion

In genetic epidemiology, multistage sampling permits the allocation of resources to families that are likely to be most informative for a given objective while still allowing population-based inference by using the sampling fractions at each stage. The only requirement for drawing valid inference is that researchers have a random sample of subjects from each stratum. Different end points may be optimized in each stage, and the selection of an optimal design depends on the aims of the study.

The potential information gain depends on the initial sample and the genetic model. Whittemore and Halpern *(12)* use simulations to show that the best sampling strategy for the case–control design for prostate cancer is to select all family history-positive case patients and a substantial fraction of family history-positive control subjects, a smaller fraction of family history-negative case patients, and the smallest fraction of family history-negative control subjects. However, our simulations for the family cancer registry design, based only on families of case patients and including the genotype of the proband, showed that the sampling weights slightly favored oversampling case patients with two or more affected relatives. Little difference was observed in the weights for sampling case patients with no affected relative and for sampling case patients with one affected relative.

Efficiency calculations for estimating gene-by-environment interaction effects from sibling data favor sampling case patients with a positive family history of disease. These same families are preferred for gene hunting with the use of linkage analysis. Combining this finding with earlier results on estimating segregation parameters resulted in a final recommendation to the Cooperative Family Registry for Colorectal Cancer Research at the University of Southern California to sample case patients with affected relatives at a higher rate than case patients without affected relatives. This procedure will improve our ability to detect gene–environment interactions and to discover new genes in latter stages of the design while maintaining the ability to characterize known genes.

Appendix A. Score Equations and Horvitz–Thompson Variance Formula

The score equation for the complete likelihood is

$$U_{ML}(\theta) = \sum_{y \in Stage2} \frac{\partial}{\partial\theta} \ln f(\theta_0;y) + \sum_{j=1}^{K} (N_{S_j} - n_{S_j}) \frac{\partial}{\partial\theta} \ln\omega_{S_j},$$

where θ denotes our parameter of interest, $f(y;\theta)$ the distribution of the stage 2 data (y), $N_{S_j} - n_{S_j}$ the number of probands classified in stratum S_j but not selected in stage 2, and ω_{S_j} the probability of being in that stratum. The Horvitz–Thompson estimating equation is

$$U_{HT}(\theta) = \sum_{j=1}^{k} \frac{1}{f_{S_j}} \sum_{y \in Stage2} \frac{\partial}{\partial\theta} \ln f(\theta_0;y),$$

where f_{S_j} denotes the sampling fraction in strata S_j. The asymptotic variance is

$$\mathrm{var}(\hat{\theta}_{HT}) = N^{-1} I(\theta_0)^{-1} + N^{-1} I(\theta_0)^{-1} B_2 (I(\theta_0)^{-1})^T,$$

where

$$I(\theta_0) = E\left[-\frac{\partial}{\partial\theta^2} \ln f(\theta_0;y) \right]$$

and

$$B_2 = \sum_{S_j} \omega_{S_j} \frac{(1 - f_{S_j})}{f_{S_j}} Var[U_{HT}(\theta_0) | Y \in S_j].$$

Appendix B. Conditional Logistic Regression

For a single case–control set, let $j = 1, \ldots, M$ index members with $j = 1$ denoting the case patient. The observed data include the environmental exposure x and the genotype g. The coding for the genotype is given by G_g. For individuals carrying two copies of the normal allele, $G_g = 0$; for those carrying two copies of the variant allele, $G_g = 1$. For subjects carrying one allele of each type, $G_g = \Delta$, where Δ is the dominance effect ($\Delta = 0$ for recessive genes, 1 for dominant genes, and ½ for genes that act multiplicatively on the relative risk scale). The parameter $\theta = (\beta, \gamma, \delta)$ denotes the logarithm of the genetic and environmental relative risks and relative risk ratio for the gene–environment interaction effect. The likelihood contribution for one family is

$$L(\theta;x,g) = \frac{\exp(\beta G_{g1} + \gamma x_1 + \delta G_{g1} x_1)}{\sum_{j=1,\ldots,M} \exp(\beta G_{gj} + \gamma x_j + \delta G_{gj} x_j)}.$$

For a sample of independent case–control sets, the likelihood is given by the product of such terms.

References

(1) Caparoso N, Rothman N, Wacholder S. Case–control studies of common alleles and environmental factors. Monogr Natl Cancer Inst 1999;26:25–30.

(2) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Monogr Natl Cancer Inst 1999;26:31–7.

(3) Gail MH, Pee D, Carroll R. Kin-cohort designs for gene characterization. Monogr Natl Cancer Inst 1999;26:55–60.

(4) Goldstein AM, Andrieu N. Detection of interaction involving identified genes: available study designs. Monogr Natl Cancer Inst 1999;26:49–54.

(5) White JE. A two-stage design for the study of the relationship between a rare exposure and a rare disease. Am J Epidemiol 1982;115:119–28.

(6) Breslow NE, Cain KC. Logistic regression for two-stage case–control data. Biometrika 1988;75:11–20.

(7) Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. Am J Epidemiol 1988;128:1198–206.

(8) Weinberg CR, Wacholder S. The design and analysis of case–control studies with biased sampling. Biometrics 1990;46:963–76.

(9) Zhao LP, Lipsitz S. Designs and analysis of two-stage studies. Stat Med 1992;11:769–82.

(10) Reilly M. Optimal sampling strategies for two-stage studies. Am J Epidemiol 1996;143:92–100.

(11) Breslow NE, Holubkov R. Weighted likelihood, pseudo-likelihood and maximum likelihood methods for logistic regression analysis of two-stage data. Stat Med 1997;16:103–16.

(12) Whittemore AS, Halpern J. Multi-stage sampling in genetic epidemiology. Stat Med 1997;16:153–67.

(13) Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite population. J Am Stat Assoc 1952;47:663–85.

(14) Flanders WD, Greenland S. Analytic methods for two-stage case–control studies and other stratified designs. Stat Med 1991;10:739–47.

(15) Pepe MS, Reilly M, Fleming T. Auxiliary outcome data and the mean score method. J Stat Planning Inference 1994;42:137–60.

(16) Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994;89:846–66.

(17) Cannings C, Thompson EA. Ascertainment in the sequential sampling of pedigrees. Clin Genet 1977;12:208–12.

Note

Supported by Public Health Service grants CA52862 (K. D. Siegmund, D. C. Thomas) and 5R35CA47448 National Institutes of Health Outstanding Investigator Grant, Research in Cancer Epidemiology and Biostatistics (A. S. Whittemore), National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

Detection of Interaction Involving Identified Genes: Available Study Designs

Alisa M. Goldstein, Nadine Andrieu

Advances in molecular genetic techniques have led to an increased ability to examine gene–environment interactions. Studies to detect gene–environment interactions are motivated by different situations, including 1) most identified cancer genes having associated lifetime risks less than 100% (i.e., incomplete penetrance), 2) hereditary factors that control the metabolism of carcinogens that may modulate risk of disease as hypothesized in pharmacogenetics, and 3) inconsistent associations across studies between a cancer and a suspected risk factor. The above situations and others have led to increased study of interaction between genetic and environmental factors. Less studied so far, but with increased potential for the future, is interaction between identified genes. Gene–gene interaction studies would also be motivated by the situations described above. Approaches to detect gene–environment and gene–gene interactions are reviewed. Available risk estimates, required types of subjects, and feasibility of the proposed study designs are discussed; efficiency and power for interaction assessment are summarized where available. In general, most designs allow for estimating risk associated with a genetic factor, environmental factor, and interaction effect. Although power and efficiency for detecting interactions have been assessed for specific situations in some of the methods, further investigations are needed to define the efficiency spectra of each design. [Monogr Natl Cancer Inst 1999;26:49–54]

Before looking at different study designs, we need to define interaction, since interaction is a model-dependent concept. We define the parameters for modeling the exposure-disease relation. Let $P(g)$ = frequency of the genetic factor g in the population and $P(e)$ = frequency of the environmental factor e in the population.

Assume OR_{eg} is the odds ratio for exposure e and genetic status g, with $e = 1$ denoting exposed and $e = 0$ denoting nonexposed and with $g = 1$ denoting the presence of the genetic factor and $g = 0$ denoting the absence of the factor. By definition, $OR_{00} = 1$.

A fundamental component of the exposure-disease relation is the specification of the joint OR, OR_{11}, under the null and alternative hypotheses. Suppose I_{00}, I_{10}, I_{01}, and I_{11} are the disease incidence rates (e.g., per person per year) for nonexposed without the genetic factor ($e = 0$ and $g = 0$), for exposed without the genetic factor ($e = 1$ and $g = 0$), for nonexposed with the genetic factor ($e = 0$ and $g = 1$), and for exposed with the genetic factor ($e = 1$ and $g = 1$), respectively. We wish to define I_{11} in terms of the other rates. One characterization of the joint association is additive, where I_{11} is the sum of the background disease rate and the excess rates for the exposure and for the genetic factor, i.e., $I_{11} = I_{00} + (I_{10} - I_{00}) + (I_{01} - I_{00}) = I_{10} + I_{01} - I_{00}$.

Dividing by I_{00}, this expression can be rewritten in terms of relative risks, as $I_{11}/I_{00} = I_{10}/I_{00} + I_{01}/I_{00} - 1$. For rare diseases, this expression is approximated by ORs, namely, $OR_{11} = OR_{10} + OR_{01} - 1$.

An alternate characterization for the joint association is multiplicative, where I_{11} is the product of the risks for the individual factors, i.e., $I_{11} = I_{10} \times I_{01}/I_{00}$. Again, dividing each side of the equation by I_{00} and assuming rare diseases, the multiplicative association is approximated by $OR_{11} = OR_{10} \times OR_{01}$. The choice of the joint association will depend on many factors, including the overall goals of the study.

Departures from multiplicativity can be assessed by the interaction effect *(1)*, which is the joint OR for the exposure and the genotype divided by the product of the ORs for the effect of exposure alone and of the genotype alone

$$\left(OR_{int.} = \frac{OR_{11}}{OR_{10}OR_{01}}\right).$$

Departures from an additive model of risk may also be assessed by dividing the joint ORs for the exposure and the genotype by the sum of the ORs for the effect of exposure alone and of the genotype alone minus 1

$$\left(OR_{int.} = \frac{OR_{11}}{OR_{10} + OR_{01} - 1}\right).$$

An interaction effect of more than one indicates a greater than multiplicative/additive effect between the exposure and the genotype, while an interaction effect of less than one indicates a less than multiplicative/additive effect.

Gene–Environment Interaction

Case-Only Study

In case-only studies, the association between an environmental exposure and a genotype is examined among case subjects only (affected subjects with a given disease). Case subjects are selected using epidemiologic principles of case selection as for any case–control study, with the exposure effect assessed among only case subjects. Case subjects without the susceptibility genotype form the control group and the nonexposed case subjects (with respect to the environmental exposure) serve as the referent group. ORs and confidence intervals are obtained using standard crude analyses or multivariate models to adjust for other covariates *(2–4)*. To yield valid estimates of interaction between

Affiliations of authors: A. M. Goldstein, Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; N. Andrieu, Unité de Recherche en Epidémiologie des Cancers, Institut de la Santé et de la Recherche Médicale U351, Institut Gustave-Roussy, 94805 Villejuif, France.

Correspondence to: Alisa M. Goldstein, Ph.D., National Institutes of Health, Executive Plaza South, MSC 7236, 6120 Executive Blvd., Bethesda, MD 20892-7236 (e-mail: (ag26o@nih.gov).

© Oxford University Press

a gene and environmental exposure, the study design requires that the exposure and genetic factor occur independently and that the disease is rare. Indeed, ORs relating exposure and the genotype among case subjects only, $OR_{case-only}$, have been shown to be a function of OR_{int} estimated from a standard case–control study and OR among control subjects relating the exposure and the genotype *(2,4)*. Under the assumption of independence of exposure and genotype, the latter OR is equal to 1. If the assumption of independence is valid and the disease is rare, the OR ($OR_{case-only}$) is the interaction effect *(4)* as measured in a traditional case–control study under a multiplicative model [for more details, *see (3)*].

Case-only studies offer better precision for estimating gene–environment interactions than those based on both case and control subjects (subjects not affected by the disease under study) (i.e., there are smaller standard errors due to elimination of control group variability) *(2)*. The power for detecting gene–environment interactions in case-only studies is comparable to the power for assessing a main effect in a classical case–control study *(5)*.

The main disadvantages of the case-only design are that the main effects of the genetic and environmental factors cannot be estimated and that the genetic and environmental factors must occur independently. In addition, even if many biologically plausible gene–environmental interaction models should cause departures from multiplicative effects *(6)*, case-only studies cannot detect gene–environmental interaction models with departures from additivity. For these reasons, Umbach and Weinberg *(7)* recently proposed an alternative method that keeps the advantages of the case-only design while simultaneously allowing estimation of the main effects. The proposed incomplete-data-case–control design collects both genotype and environmental exposure data from the case subjects but only environmental exposure or only genotype data from the control subjects. The estimation of main effect(s) is possible only if the genotype and environmental exposures are independent and the studied disease is rare. Umbach and Weinberg *(7)* proposed a maximum likelihood method based on log-linear models, which allows imposition of the independence assumption, whereas usual logistic regression does not permit such an imposition. The authors showed that their method may need fewer than half as many individuals as methods that do not impose the gene–environment independence assumption to reach the same power for detecting gene–environment interaction. This approach allows estimation of interaction that is a departure from either multiplicativity or additivity in the relative risk. However, in case-only and incomplete-data- case–control designs, the required assumption of independence between a gene and environmental factor may be violated when the exposure and the gene both vary with an unmeasured variable. To ensure the validity of the critical independence assumption, a random sample of control subjects with both genetic and exposure data should be collected. This additional requirement may lead to a questionable gain in efficiency for this design compared with a full case–control design.

Case–Control Study Design Using Unrelated Control Subjects

The case–control study with population-based, hospital-based, or other unrelated control subjects is the most commonly used design in studies of gene–environment interaction. By use of unexposed subjects with no susceptibility genotype as the referent group, ORs for all other groups can be estimated under either multiplicative or additive models. Adjustment for potential confounding variables may be accomplished by using stratification or multivariate approaches.

The power of this traditional case–control study design for assessing interaction was first studied examining interaction between unspecified risk factors *(1,8,9)* and then focusing on gene–environment interaction *(10–12)*. Among the parameters needed to determine the power to detect a gene–environment interaction, the frequencies of the environmental exposure and the genetic factor appeared to be the most important in determining the number of case subjects required. In fact, results from multiple studies *(10–12)* suggest that case–control designs may be used to detect gene–environment interaction only when the environmental factor and genetic factor are common. Indeed, certain combinations of ORs and frequencies of the genetic factor and the environmental factor make study sizes prohibitive.

The main advantage of this design is that the main effects of the environmental exposure and genetic susceptibility as well as their interactive effect may be estimated. The main disadvantages are that this design may not be appropriate for the study of gene–environment interaction involving rare genes or uncommon environmental exposures (assuming reasonable values of the interaction effect). In addition, population stratification or genetic admixture might adversely influence this design. The extent of the problem produced by population stratification is discussed elsewhere *(13)*.

Two-Stage Case–Control Study: Countermatching and Balanced Design

To detect a gene–environment interaction when one of the factors under study is rare, one approach would be to conduct a two-stage or multistage design that increases the numbers of case and control subjects with the rare factor of interest without prohibitively increasing the number of measurements to perform *(14)*.

For two-stage study designs, samples of case and control subjects are drawn from the population at risk. After classification according to an exposure of interest, subsamples of case and control subjects are selected for purposes of covariate assessment. Approaches include balancing the numbers of exposed and unexposed case and control subjects as well as countermatching in which control subjects are countermatched rather than matched to case subjects. The goal of both approaches is to improve the statistical efficiency of exposure risk estimates compared with the classical random case–control study *(15,16)*.

Countermatching designs. Countermatching has been recently proposed by Langholz and Clayton *(16)* as a method of sampling controls from a cohort for nested case–control studies. In gene–environment interaction assessment, one scenario for countermatching would be to use a surrogate of the disease gene such as family history, which is generally available on the entire cohort/population at risk. In contrast, a measure of the gene itself would likely be too costly to obtain for the whole cohort. Assuming the disease is rare, all case subjects from the cohort/population at risk would be sampled. Each case subject's risk set would be stratified by the surrogate of the gene, i.e., family history of disease, and control subjects for that risk set would be selected from the strata other than the case subject's stratum. The goal of countermatching is to maximize the number of discordant case–control pairs from which information comes in

a matched case–control study. The gain in efficiency will depend on how predictive family history of disease is for the gene of interest. Alternatively, one could countermatch on a rare environmental exposure or on both the environmental factor and the genetic one. A partial likelihood has been developed to estimate exposure effects in countermatching *(17)* using weighting that takes into account the probability that subjects were selected from specific strata. This method has been shown to increase the efficiency of main effect estimation by approximately 25% compared with classical random sampling *(18)*. In gene–environment interaction assessment, countermatching on both the gene and environmental factor has been shown to be more efficient than a standard nested case–control study using three control subjects per case as well as designs that countermatched on either the environmental factor or the gene. The relative efficiencies were influenced mainly by the frequencies of the true risk factors and the sensitivity and specificity of the factors' surrogates (Andrieu N, Goldstein AM, Thomas DC, Langholz B: manuscript submitted for publication). Studies to examine additional scenarios and evaluate cost, efficiency, and feasibility need to be conducted to determine the utility of these approaches for interaction assessment.

Balanced designs. A more intuitive approach to studying a rare factor would be to oversample for the rare covariate of interest. If the gene or environmental factor under study is rare, rather than choose a subset at random, case and control subjects may be sampled nonrandomly, that is oversampled for the rare factor (called exposure). Thus, for example, stage II of a two-stage design would consist of selecting all case subjects (if the disease is rare) exposed to the rare factor and sampling an equal number in each of the three other categories. Then the other factor (called covariate) involved in the gene–environment interaction is measured in this subsample. The oversampling is taken into account in the analysis to obtain unbiased estimates of the exposure effect and interaction effect in which the exposure is involved *(15,19)*. Cain and Breslow *(19)* investigated the efficiency of a balanced design compared with a random sampling case–control design, particularly for estimating exposure-covariate interaction. Given a rare exposure (prevalence = 0.05; OR = 2.0) and a covariate of moderate frequency (0.3), with a range of ORs and correlations between the exposure and covariate as measured by the ORs in the control group, the results showed that the balanced design was always much more efficient than a random sampling design for estimating the exposure-covariate interaction in terms of standard errors for both a rare and a common disease. The efficiency decreased as the correlation between the exposure and covariate increased but still remained substantial. For example, when the degree of confounding between the exposure and covariate was set at 5.0, the relative efficiency of the balanced design was greater than three for a common disease and equal to two for a rare disease.

In the two-stage designs, the rare exposure is assumed to be known for the entire population at stage I. Thus, if a rare genetic or environmental factor requires expensive investigation to be measured, measurement may not be affordable at the first stage. One could then propose a three-stage study with two successive sampling stages. Stage I would consist of selecting all case subjects (if the disease is rare) exposed to an inexpensive surrogate variable for the factor under study and sampling equal numbers of unexposed case subjects and exposed and unexposed control subjects. The factor assessment could then be performed on the stage II subsample. Stage III would consist of selecting all case subjects exposed to the rare factor and sampling an equal number of unexposed case subjects and exposed and unexposed control subjects. The environmental or genetic factor of interest for the gene–environment interaction evaluation would be collected in this stage III subsample. Such an approach needs to be statistically evaluated before being considered. At present, identifying good surrogates for the factor(s) of interest and the costs associated with measuring a genetic or environmental factor on large numbers of subjects may be the major determinants for deciding whether or not to conduct multistage balanced and countermatched designs.

Case–Control Study Design Using Related Control Subjects

To our knowledge, a case–control study design using related control subjects has not yet been applied in studies of gene–environment interaction. In this approach, each case subject is matched to one or more unaffected relative(s) and a standard conditional logistic regression is performed to assess gene–environment interaction. This method does not result in biased estimates and does yield consistent estimates, even when there is a correlation in risk factors under study between relatives *(20,21)*. However, bias may occur if there is a correlation, within matched case-control pairs, between measured and unmeasured risk factors *(22,23)*.

Several variations of this study design have been proposed. The use of unaffected siblings has been proposed to study the role of environmental factors and specific gene loci and to provide evidence of gene–environment interaction *(24)*. Case subjects are affected siblings of the proband (case subjects who led to family selection), and control subjects are randomly chosen from unaffected siblings of the proband. The genetic factor (e.g., exposure) for each case and control subject is defined as the number of alleles shared identical by descent (ibd) with the proband at the studied locus. Disease risks can then be estimated for the three groups: share 0, share 1, and share 2. Under the null hypothesis of no linkage between the disease susceptibility locus and the genetic factor locus, disease risks are expected to be identical in the three groups. ORs associated with the genetic factor can be obtained, for example, by comparing the share-2 and share-1 allele groups with the baseline risk group, share-0 alleles. Covariates including environmental exposures may be investigated for main effects and for evidence of interaction with the studied locus by using stratification or multivariate analyses. An alternative matched design has also been proposed for matching on family. One form of this design requires families with at least one affected and one unaffected sibling in addition to the proband limiting the number of available families. An alternative form uses all family members from eligible families in a conditional logistic regression approach. Further methodologic studies are needed to examine issues related to design, conduct, and analysis of these types of studies. In addition, power, efficiency, and feasibility for gene–environment interaction detection using these approaches have not yet been thoroughly investigated.

Another design is based on the use of parental genotypes *(25,26)*. One method of analysis (transmission disequilibrium test [TDT] method) consists of comparing the genotype of each case subject with the genotype of a fictitious control subject composed of the nontransmitted alleles from each parent. The

method was first proposed for genetic markers-disease association studies. It has been extended by stratifying case subjects according to the presence or absence of an environmental exposure. The ORs associated with the genetic factor can be obtained for nonexposed and exposed case subjects. Differences in the estimates across strata may reflect gene–environment interaction *(27)*. One disadvantage of this study design is that it does not allow for assessing the main effect of the environmental exposure. Also, differences in estimates across strata are not restricted to resulting from only gene–environment interaction. Differentiating between the various reasons for the differences may be difficult.

The advantage common to these case-related control study designs is the elimination of control subjects whose genetic backgrounds differ systematically from those of case subjects by using relatives as the control subjects. In addition, relatives are easily identified and may be more willing to participate in a research study. Some of these approaches permit estimation of the main effects of the environmental exposures and genetic factors as well as their interaction effects, whereas other approaches only allow estimation of the interaction effect. There may also be a gain in efficiency for detecting gene–environment interaction involving a rare gene compared with a traditional case–control study approach because of oversampling on the genetic factor by using related control subjects. Witte et al. *(28)* evaluated the efficiency of using either sibling or cousin control subjects for assessing the main effect of the genetic factor and the gene–environment interaction effect. Although relative control subjects were less efficient than population-based control subjects for detecting the genetic factor main effect, sibling control subjects were the most efficient control group for detecting interaction. The gain in efficiency, however, decreased as the frequency of the genetic factor increased *(28)*. One disadvantage of these case-related control designs is the potential for overmatching, particularly on environmental exposures of interest leading to a loss in efficiency. The extent of this loss will depend on the exposures of interest and their correlation among relatives. Thus, there is a potential paradox—efficiency may be improved by oversampling; it may also be reduced by oversampling. The relative efficiency will depend on a number of parameters, including the environmental exposure and genetic factor concordance between relatives. Another possible disadvantage is the cost of finding and ascertaining the relative control subjects. In addition, for designs requiring known ibd status (e.g., comparisons of affected and unaffected siblings), parents must be genotyped otherwise ibd status will not be known and identity by state will need to be used. Also, there may be additional ethical issues associated with using related control subjects versus unrelated control subjects.

Family Study Design

Another proposed study design is a family study with no separate control group. Families are recruited through one or more diseased members, and data on disease status and covariates are collected on all family members. The proposed analysis is segregation analysis or combined segregation and linkage analysis (without or with a known candidate gene), the goals of which are to assess gene–environment interaction. Recently, several approaches have been developed to allow direct assessment of gene–environment interaction *(29–33)*. The study designs allow for estimating genetic, environmental, and gene–environment interaction effects using somewhat different approaches. The regressive models *(31)* were constructed by specifying a relationship between each person's phenotype for a studied trait and a set of explanatory variables, including the person's major genotype, the phenotype of relatives to take into account residual family dependencies of unspecified origin (genetic or environmental), and measured covariates. Studied traits were able to be either quantitative or qualitative. The major gene is assumed to be in Hardy–Weinberg equilibrium following the rules of mendelian transmission through generations. Abel and Bonney *(34)* introduced survival analysis concepts into the regressive models to take into account variable ages at trait diagnosis. The censored survival model *(33)* is an extension to the Cox proportional hazards model and includes measured covariates and unobserved covariates for the genetic component adjusting for variable age at diagnosis for the trait. Two unobserved latent variables are considered: a diallelic major gene and a polygene. The difference between the two approaches in terms of explanatory variables is in residual family likeness for the studied trait that can be taken into account by a polygene in Gauderman and Thomas' censored survival model and by family dependencies in the original model of Bonney. As such, an extra interaction between the environment factor(s) and the polygene may be taken into account in the censored survival approach.

To date, these approaches have been rarely used to detect gene–environment interaction in human diseases. The major limitations are the need to collect individual data on the trait of interest and environmental covariate data on all family members in a sample size sufficient to have power to detect an interaction. Few investigators have been able to collect such data. In addition, the effects of differential participation of family members related to survival, availability, and motivation needs to be assessed. Although limited examinations of power and efficiency for the censored survival model have been conducted *(33,35,36)*, further methodologic studies are needed to evaluate a larger spectrum of models by varying the mode of inheritance, allele frequencies, and relative risks associated with the environmental factor(s), major gene, and interaction effect. Similarly, further methodologic studies are needed to evaluate regressive model efficiency in gene–environment interaction detection for different interaction patterns.

Partly because of some of the difficulties in obtaining genetic and environmental factor data on multiple family members, modifications of the above approaches have been proposed. These variations include family-based cohort designs (e.g., the kin-cohort design) *(37)*, family-based case–control association studies *(38)*, and case–control family study designs *(39)*. These approaches are discussed further elsewhere in the Workshop.

Gene–Gene Interaction

All of the study designs previously described, i.e., case-only study, incomplete-data-case–control study, case–control study using unrelated or related control subjects, and multistage case–control study designs, are adaptable for gene–gene interaction assessment. Indeed, the approach is essentially identical and one may replace the environmental factor by a second genetic factor. In family study designs, a second genetic factor would be introduced in segregation and joint segregation and linkage analyses as a covariate similar to gene–environment assessment. In the case where one or both of the two genetic factors are not yet identified, two-locus segregation, linkage, and linkage/associa-

Table 1. Study designs for gene–environment interaction assessment

Study design (reference No.)*	Risk estimate available for			Exposure information needed for control subjects		
	Genetic factor	Environmental factor	Interaction	Relative control subjects		Unrelated control subjects
				Affected	Unaffected	
Case only *(2–4)*	No	No	Yes	—	—	—
Incomplete-data-case–control *(7)*	Yes†	Yes†	Yes	—	—	Yes†
Case–control						
Unrelated control subjects *(10–12)*	Yes	Yes	Yes	—	—	Yes
Two stage [balanced *(15,19)*, countermatching *(16)*]	Yes	Yes	Yes	—	—	Yes
Related control subjects *(28)*	Yes	Yes	Yes	—	Yes	—
Case-unaffected-sibling pair *(24)*	Yes	Yes	Yes	Yes	Yes	—
Case-parental control (or TDT or case-pseudosibling design) *(25–27)*	Yes	No	Yes‡	—	Yes	—
Family study *(28–35)*	Yes	Yes	Yes	Yes	Yes	—

*TDT = transmission disequilibrium test.

†Only environmental exposure or only genotype data collected on control subjects. Possible to assess main effect(s) if genotype or environmental exposure data collected on control subjects.

‡Differences across strata not restricted to resulting from only gene–environment interaction.

tion methods may be used. These methods include the MASC method *(40)*, the COMDS method *(41)*, parametric *(42)*, and nonparametric *(43)* two-locus linkage analyses. This topic will not be further discussed here. Examination of two-locus segregation, linkage, and linkage/association methods are presented by the Gene Discovery Panel *(44)*.

Conclusions

Table 1 presents the study designs reviewed here and the references for the designs. Most designs currently available for examining interactions allow for estimating risk associated with a genetic factor, environmental exposure factor, and interaction effect. Case-only studies and TDT-like approaches, however, allow for assessing interaction effects only. Since case-only studies can efficiently estimate gene–environment interactions, though, one might consider using this approach if resources are sufficient to only genotype cases, if one is interested in studying the effect of gene–environment interactions on tumor characteristics, if only tumor specimens are available, or if there are institutional review board concerns about linking genetic and environmental exposure data in control subjects. On the other hand, if the exposure risk is not already well known, which may be the case when the genetic factor involved in the interaction is common, the usefulness of case-only study designs may be questionable and the incomplete-data-case–control design developed by Umbach and Weinberg *(7)* might be recommended as an alternative. Assessment of gene–environment interaction effects without estimating and understanding the main effects of the genetic and environmental factors would be of little use for public health or individual risk assessment. However, there is potential for increasing knowledge about biologic mechanisms underlying gene–environment interaction. The critical assumption for these studies, the independence of the genetic and environmental factors, requires linked genetic and environmental factor data in at least some control subjects *(7)*.

Efficiency and power for gene–environment and gene–gene interaction assessment have been rarely studied for the various designs presented, and further investigations are needed to define the efficiency spectra of each approach in interaction assessment. Most available designs seem to be inefficient for detecting interaction of a rare factor(s). Multistage balanced or countermatched studies could be an alternative approach if the rare event is easily and inexpensively measured. As such, at present, it is likely too costly to use this approach to study rare genetic factors. One cautionary note should be added. As the number of known genes increases, the ability to assess gene–environment interaction will also increase. This multiple testing may lead to false evidence for interaction; thus, care must be exercised when evidence for a new gene–environment interaction is first discovered. As is true for new observations in general, confirmation of a new gene–environment interaction as well as a new main effect should be sought.

To date, no single design appears universally applicable for assessing gene–environment or gene–gene interactions. The most appropriate approach(es) will depend on the disease, environmental exposure, genetic factor(s) and the interaction effect, their risk values and frequencies, and how much data are available on each of these components. Further assessment of existing designs and development of new approaches are needed to improve the ability to detect gene–environment and gene–gene interactions in complex diseases.

References

(1) Smith PG, Day NE. The design of case–control studies: the influence of confounding and interaction effects. Int J Epidemiol 1984;13:356–65.

(2) Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case–control studies. Stat Med 1994;13:153–62.

(3) Begg CB, Zhang ZF. Statistical analysis of molecular epidemiology studies employing case-series. Cancer Epidemiol Biomarkers Prev 1994;3:173–5.

(4) Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene–environment interaction: case–control studies with no controls. Am J Epidemiol 1996;144:207–13.

(5) Breslow NE, Day NE. The analysis of case–control studies. In: Davis W, editor. Statistical methods in cancer research. Lyon (France): IARC; 1980.

(6) Khoury MJ, Adams MJ, Flanders WD. An epidemiologic approach to ecogenetics. Am J Hum Genet 1988;42:89–95.

(7) Umbach D, Weinberg C. Designing and analysing case–control studies to exploit independence of genotype and exposure. Stat Med 1997;16:1731–43.

(8) Greenland S. Test for interaction in epidemiologic studies: a review and a study of power. Stat Med 1983;2:243–51.
(9) Thomas DC, Greenland S. The efficiency of matching in case–control studies of risk factor interactions. J Chronic Dis 1985;38:569–74.
(10) Hwang SJ, Beaty TH, Liang KY, Coresh J, Khoury MJ. Minimum sample size estimation to detect gene–environment interaction in case–control designs. Am J Epidemiol 1994;140:1029–37.
(11) Goldstein AM, Falk RT, Korczak JF, Lubin JH. Detecting gene–environment interactions using a case–control design. Genet Epidemiol 1997;14: 1085–9.
(12) Foppa I, Spiegelman D. Power and sample size calculations for case–control studies of gene–environment interactions with a polytomous exposure variable. Am J Epidemiol 1997;146:596–604.
(13) Caporaso N, Rothman N, Wacholder S, Case–control studies of common alleles and environmental factors. Monogr Natl Cancer Inst 1999;26:25–30.
(14) Siegmund KD, Whittemore AS, Thomas DC. Multistage sampling for disease family registries. Monogr Natl Cancer Inst 1999;26:43–8.
(15) Breslow NE. Two phase case–control studies. In: Armitage P, Colton T, editors. Encyclopedia of biostatistics. New York (NY): John Wiley & Sons; 1998;1:532–40.
(16) Langholz B, Clayton D. Sampling strategies in nested case–control studies. Environ Health Perspect 1994;102(suppl 8):47–51.
(17) Langholz B, Borgan O. Counter-matching: a stratified nested case–control sampling method. Biometrika 1995;82:69–79.
(18) Steenland K, Deddens JA. Increased precision using countermatching in nested case–control studies. Epidemiology 1997;8:238–42.
(19) Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. Am J Epidemiol 1988;128:1198–206.
(20) Goldstein AM, Hodge SE, Haile RW. Selection bias in case–control studies using relatives as the controls. Int J Epidemiol 1989;18:985–9.
(21) Robins J, Pike M. The validity of case–control studies with non random selection of controls. Epidemiology 1990;1:273–84.
(22) Gladen BC. Matched-pair case–control studies when risk factors are correlated within the pairs. Int J Epidemiol 1996;25:420–5.
(23) Andrieu N, Goldstein AM. Use of relatives of cases as controls to identify risk factors when an interaction between environmental and genetic factors exists. Int J Epidemiol 1996;25:649–57.
(24) Khoury MJ, Flanders WD, Lipton RB, Dorman JS. The affected sib-pair method in the context of an epidemiologic study design. Genet Epidemiol 1991;8:277–82.
(25) Ewens WJ, Spielman RJ. The transmission/disequilibrium test: history, subdivision and admixture. Am J Hum Genet 1995;57:455–64.
(26) Khoury MJ, James LM. Population and familial relative risks of disease associated with environmental factors in the presence of gene–environment interaction. Am J Epidemiol 1992;137:1241–50.
(27) Maestri NE, Beaty TH, Hetmanski J, Smith EA, McIntosh I, Wyszynski DF, et al. Application of transmission disequilibrium tests to nonsyndromic oral clefts: including candidate genes and environmental exposures in the models. Am J Med Genet 1997;73:337–44.
(28) Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case–control studies of candidate genes and gene–environment interactions: basic family designs. Am J Epidemiol 1999;149:693–705.
(29) Moll PP, Sing CF, Lussier-Cacan S, Davignon J. An application of a model for a genotype-dependent relationship between a concomitant (age) and a quantitative trait (LDL cholesterol) in pedigree data. Genet Epidemiol 1984;1:301–14.
(30) Konigsberg LW, Blangero J, Kammerer CM, Mott GE. Mixed model segregation analysis of LDL-C concentration with genotype-covariate interaction. Genet Epidemiol 1991;8:69–80
(31) Bonney GE. Regression logistic models for familial disease and other binary traits. Biometrics 1986;42:611–25
(32) Demenais F, Lathrop M. REGRESS: a computer program including the regressive approach into the LINKAGE programs. Genet Epidemiol 1994; 11:291.
(33) Gauderman WJ, Thomas DC. Censored survival models for genetic epidemiology: a gibbs sampling approach. Genet Epidemiol 1994;11:171–88.
(34) Abel L, Bonney G. A time-dependent logistic hazard function for modeling variable age of onset in analysis of familial diseases. Genet Epidemiol 1990;7:391–407.
(35) Gauderman WJ, Morrison JL, Carpenter CL, Thomas DC. Analysis of gene–smoking interaction in lung cancer. Genet Epidemiol 1997;14: 199–214.
(36) Gauderman WJ, Faucett CL. Detecting gene–environment interactions in joint segregation and linkage analysis. Am J Hum Genet 1997;61:1189–99.
(37) Gail MH, Pee D, Carroll R. Kin-cohort designs for gene characterization. Monogr Natl Cancer Inst 1999;26:55–60.
(38) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Monogr Natl Cancer Inst 1999;26:31–7.
(39) Zhao LP, Aragaki C, Hsu L, Potter J, Elston R, Malone KE, et al. Integrated designs for gene discovery and characterization. Monogr Natl Cancer Inst 1999;26:71–80.
(40) Dizier MH. Babron MC. Clerget-Darpoux F. Interactive effect of two candidate genes in a disease: extension of the marker-association-segregation chi(2) method. Am J Human Genet 1994;55:1042–9.
(41) Morton NE, Shields DC, Collins A. Genetic epidemiology of complex phenotypes. Ann Hum Genet 1991;55:301–14.
(42) Schork NJ, Boehnke M, Terwilliger JD, Ott J. Two-trait-locus linkage analysis: a powerful strategy for mapping complex genetic traits. Am J Hum Genet 1993;53:1127–36.
(43) Knapp M, Seuchter SA, Baur MP. Two-locus disease models with two marker loci: the power of affected-sib-pair tests. Am J Hum Genet 1994; 55:1030–41.
(44) Schaid DJ, Buetow K, Weeks DE, Wijsman E, Guo SW, Ott J, et al. Discovery of cancer susceptibility genes: study designs, analytic approaches, and trends in technology. Monogr Natl Cancer Inst 1999;26: 1–16.

Kin-Cohort Designs for Gene Characterization

Mitchell H. Gail, David Pee, Raymond Carroll

Background: **In the kin-cohort design, a volunteer with or without disease (the proband) agrees to be genotyped, and one obtains information on the history of a disease in first-degree relatives of the proband. From these data, one can estimate the penetrance of an autosomal dominant gene, and this technique has been used to estimate the probability that Ashkenazi Jewish women with specific mutations of BRCA1 or BRCA2 will develop breast cancer.** ***Methods:*** **We review the advantages and disadvantages of the kin-cohort design and focus on dichotomous outcomes, although a few results on time-to-disease onset are presented. We also examine the effects of violations of assumptions on estimates of penetrance. We consider selection bias from preferential sampling of probands with heavily affected families, misclassification of the disease status of relatives, violation of Hardy–Weinberg equilibrium, violation of the assumption that family members' phenotypes are conditionally independent given their genotypes, and samples that are too small to ensure validity of asymptotic methods.** ***Results and Conclusions:*** **The kin-cohort design has several practical advantages, including comparatively rapid execution, modest reductions in required sample sizes compared with cohort or case–control designs, and the ability to study the effects of an autosomal dominant mutation on several disease outcomes. The design is, however, subject to several biases, including the following: selection bias that arises if a proband's tendency to participate depends on the disease status of relatives, information bias from inability of the proband to recall the disease histories of relatives accurately, and biases that arise in the analysis if the conditional independence assumption is invalid or if samples are too small to justify standard asymptotic approaches. [Monogr Natl Cancer Inst 1999;26: 55–60]**

Wacholder et al. *(1)* used the term "kin-cohort" design to describe a study in which volunteers (probands) are genotyped and one also determines the phenotypes of probands' first-degree relatives. This design was employed by Struewing et al. *(2)* to estimate the cumulative probability of developing breast cancer, as a function of age, for carriers of mutations of BRCA1 or BRCA2 in Ashkenazi Jews from the region surrounding Washington, DC. That study indicated that the cumulative risk of breast cancer to age 70 years was 0.56, a number that was substantially lower than estimates from highly affected pedigrees *(3)*, but not inconsistent with the estimate of 0.69 obtained by Whittemore et al. *(4)* from segregation analyses (without any genotyping) of U.S. population-based case subjects and control subjects and their first-degree relatives. Although the term "kin-cohort design" might be used for any study of a family that is selected on the basis of a proband, we shall reserve its use to those studies in which the probands are genotyped.

For population-based inference, Gail et al. *(5)* stressed the importance of representative sampling of probands, conditional on their phenotypes. Thus, for dichotomous phenotypes with $Y_0 = 1$ denoting a case proband and $Y_0 = 0$ denoting a control proband, Gail et al. assumed that the case probands were selected at random from among all case subjects in the population and the control probands were selected at random from among all control subjects in the population. One can enrich the frequency of mutation carriers in a kin-cohort study by increasing the proportion of case probands.

Wacholder et al. *(1)* estimated the cumulative probability of developing breast cancer to a given age by recognizing that the cumulative incidence distribution for first-degree relatives of probands who carried the mutation was a mixture (nearly 50 : 50) of carrier and noncarrier distributions, whereas the cumulative incidence distribution for first-degree relatives of probands who did not carry the mutation was a different mixture (almost 0 : 100 of carrier and noncarrier distributions). It was thus possible to estimate the underlying carrier and noncarrier distributions from data on the cumulative incidence in relatives of carrier and noncarrier probands. Gail et al. *(5)* estimated penetrance using likelihood methods similar to those in segregation analysis but adapted to take into account the genotypes of the probands.

The kin-cohort design can be extended by genotyping some of the relatives of the proband *(5)* or by expanding the information on relatives to include the phenotypes of second-degree and more distant relatives. Gail et al. *(5)* used the term "genotyped-proband design" to stress that the proband is genotyped in the kin-cohort design. Studies of first-degree relatives of case and control probands in the absence of genotyping, such as those described by Whittemore et al. *(4)* and by Claus et al. *(6)*, have been called "case–control family studies." These studies can be analyzed with the use of classical methods of segregation analysis. Thomas *(7)* used the term "family cohort study" whether or not the proband is genotyped.

There are several practical advantages of the kin-cohort design that invite its use for estimating penetrance *(1)*. Often one can field and complete the study quickly. Because the disease status of the relatives can be determined at the date of recruitment of the proband, one does not need to wait to acquire such information as in a prospective cohort study. In some populations and for some diseases, one may obtain reliable information on the phenotypes of first-degree relatives simply by interviewing the proband. In such circumstances, time is saved because a single proband interview yields the phenotypes of several relatives. Moreover, one can inquire about more than one disease in the same proband interview. For example, Struewing et al. *(2)* demonstrated an increased risk for ovarian and prostate cancers among carriers of BRCA1 and BRCA2 mutations as well as an

Affiliations of authors: M. H. Gail, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; D. Pee, Information Management Services, Rockville, MD; R. Carroll, Texas A&M University, College Station.

Correspondence to: Mitchell H. Gail, M.D., Ph.D., National Cancer Institute, EPS-8032, Bethesda, MD 20892 (e-mail: gailm@exchange.nih.gov).

See "Notes" following "References."

© Oxford University Press

increased risk of breast cancer. One can also study the penetrances associated with additional mutations simply by genotyping the proband for the several mutations. Furthermore, as shown by Gail et al. *(5)*, the sample sizes required by the kin-cohort design to estimate penetrance of an autosomal dominant mutation with a desired degree of precision are comparable to or somewhat smaller than the sample sizes needed for a cohort study or for a population-based case–control design and can be substantially smaller if some relatives are also genotyped.

Estimates of penetrance from the kin-cohort design are, however, subject to certain biases that arise when assumptions underlying the analysis are not met. In this article, we review the assumptions underlying a likelihood-based analysis of the kin-cohort design (*see* "Methods" section) and give an example of sample sizes required by this design, compared with cohort or case–control designs (*see* "Results" section). Then we discuss the effects of violations of the assumptions on estimates of penetrance and allele frequency (*see* "Results" section). Finally, we consider some conditions favorable for the use of the kin-cohort design and some potential applications apart from estimating penetrance (*see* "Discussion" section).

Methods

Several assumptions underlie a standard likelihood analysis of kin-cohort data, as described by Gail et al. *(5)*. For this article, we usually assume that phenotype is dichotomous, although methods for time-to-disease onset data have also been presented by Wacholder et al. *(1)* and Gail et al. *(5)*. A few simulations and analyses are based on parametric methods for survival data given by Gail et al. *(5)*, and we comment further on survival methods in the "Results" section. The key assumptions used in this article, some of which can be relaxed, are as follows: A1) Risk follows an autosomal dominant pattern, in which carriers of the mutant allele have a chance of disease (penetrance) ϕ_1 and noncarriers have penetrance ϕ_0; A2) the mutant allele "*A*" and wild-type "*a*" are in Hardy–Weinberg equilibrium (HWE); A3) given a family member's genotype, his or her phenotype is conditionally independent of the phenotypes of other family members; A4) probands with phenotype Y_0 are selected at random from members of the population with phenotype Y_0; A5) disease status is determined without error; and A6) sample sizes are large enough to justify standard asymptotic methods.

Under these assumptions, the likelihood for a typical family selected on the basis of a proband with phenotype Y_0 is

$$P(Y_1 \mid g_0)P(g_0) \mid Y_0), \qquad [1]$$

where $g_0 = 1$ or 0 according as the proband is a carrier or not and Y_1 is a vector of phenotype indicators, one for each relative. The use of the conditional probability $P(g_0 \mid Y_0)$, instead of $P(g_0, Y_0)$, allows one to preferentially select diseased probands and is justified by assumption A4. The term $P(Y_1 \mid g_0)$, instead of $P(Y_1 \mid g_0, Y_0)$, is based on the conditional independence assumption A3, which also leads to

$$P(Y_1 \mid g_0) = \Sigma P(Y_{11} \mid g_{11})P(Y_{12} \mid g_{12}) \ldots P(Y_{1m} \mid g_{1m})P(g_1 \mid g_0), \qquad [2]$$

where the summation is over the vector g_1 of indicators of the m relatives' carrier statuses. In equation 2, $P(Y_{1i} \mid g_{1i})$ depends only on ϕ_1 for $g_{1i} = 1$ or ϕ_0 for $g_{1i} = 0$. The probability mass function $P(g_1 \mid g_0)$ depends only on the allele frequency $P(A) = q$ and is calculated by standard mendelian methods under HWE. If relatives were also genotyped, the first term in equation 1 would become the product of $P(g_1 \mid g_0)$, which depends only on q, and of $P(Y_1 \mid g_1)$, which depends only on ϕ_1 or ϕ_0. Unless otherwise noted, we shall assume that only the phenotypes of $m = 2$ relatives are available for each proband. Further computational details are available in Gail et al. *(5)* and in a companion paper *(8)*.

Results

Sample Size

The following example is based on estimates by Claus et al. *(6)* of the penetrance of alleles predisposing to breast cancer. Adapting their results for dichotomous outcomes, Gail et al. *(5)* used $q = 0.0033$, $\phi_1 = 0.92$, and $\phi_0 = 0.10$ to represent a rare autosomal dominant allele with high penetrance.

In order to estimate ϕ_1 with precision $\pm 1.96\{\mathrm{Var}(\hat{\phi}_1)\}^{1/2} = \pm 0.05$, a cohort study would require 114 mutation carriers. To identify these carriers, however, one would expect to need to screen 17301 subjects because the allele is so rare. This might be possible in the context of a retrospective cohort study in which biologic materials were stored on a large number of women, but the effort would be substantial. A prospective cohort study would require many years of follow-up after the massive screening effort.

It is also possible to estimate ϕ_1 and ϕ_0 from a population-based case–control study if the probability of disease $P(Y_0)$ is known [or if the allele frequency is known, as discussed by Gail et al. *(5)*]. Gail et al. *(5)* found that 17030 case subjects and control subjects would need to be genotyped to achieve the desired precision on $\hat{\phi}_1$ using the optimal case : control ratio, 1524 : 15506.

For comparison, a kin-cohort design in which 10% of the probands are cases and 90% are controls requires genotyping 16080 probands if there are $m = 1$ relatives per proband. If there are $m = 2$ relatives per proband, 14935 probands are required *(5)*. Surprisingly, using only case probands does not decrease the required sample sizes necessarily. With one relative per proband, 26851 case probands are needed, instead of the 16080 probands above, while with two relatives per proband, 13418 case probands are needed instead of 14935.

Substantial reductions in the numbers of required genotypes can be obtained in the kin-cohort design if it is possible to genotype the relatives. For example, with $m = 1$ and with all case probands, if one genotypes both the proband and the relative, then 3549 families are required and $2 \times 3549 = 7098$ genotypes. If one only genotypes the relative when the proband is a carrier, then 3940 families and 4167 genotypes are required *(5)*.

Thus, the kin-cohort design offers only modest reductions in sample size, compared with cohort or case–control designs, unless it is also possible to genotype relatives. Siegmund et al. *(9)* describe increases in efficiency from genotyping additional family members in two-stage designs.

As the allele frequency increases, the required sample sizes for the kin-cohort design decrease dramatically, as is also the case for cohort and case–control designs. Gail et al. *(5)* give tables that allow one to estimate sample sizes required to obtain required precision on estimates of ϕ_1. These tables cover a range of allele frequencies, allow for $m = 1$ or 2 relatives per proband, and allow for the possibility of genotyping a relative.

Violations of Assumptions

Selection bias. A serious bias can arise if probands agree to participate more readily if they have affected relatives, violating assumption A4. Gail et al. *(8)* examined families with a mother proband and $m = 2$ daughter relatives and with $\phi_1 = 0.9$, $\phi_0 = 0.1$, and $q = 0.1$. They assumed that a mother with one affected daughter was twice as likely to participate as a mother with no affected daughters and that a mother with two affected daughters was four times as likely to participate, regardless of the mother's phenotype. (We call this "1:2:4 selection bias.") They used designs with 1154 families to achieve precision ±5% on $\hat{\phi}_1$. In simulations with 5000 repetitions, they found mean estimates (with standard errors) of 0.944 (0.009) for $\hat{\phi}_1$, 0.150 (0.015) for $\hat{\phi}_0$, and 0.205 (0.015) for $\hat{q}$. Thus, both the penetrances and allele frequency were seriously overestimated.

Similar overestimates were found when parameter values $\phi_1 = 0.56$, $\phi_0 = 0.13$, and $q = 0.0116$ were chosen to reflect the data in Struewing et al. *(2)*. Because ϕ_1 is nearer 0.5 than in the previous example, we required more families (i.e., 7634) to achieve the required precision, and half of these families had case probands. The average estimates (with standard errors) were 0.690 (0.021), 0.214 (0.003), and 0.018 (0.001), respectively, based on 559 simulations. Thus, all the parameters were overestimated, as in the previous example.

Cannings and Thompson *(10)* commented on similar ascertainment biases in segregation analyses that arise when families are selected partly on the basis of relatives' phenotypes. Siegmund et al. *(9)* reviewed methods for systematically oversampling probands with affected relatives. When such oversampling is designed, analytical corrections are available to eliminate bias; however, in usual kin-cohort designs, one has no control over who volunteers. Even in kin-cohort designs in which one can select representative samples of probands, one will not control which selected probands agree to participate or know why some refuse.

Misclassification of relatives' phenotypes. If one relies on the proband to provide information on relatives' phenotypes, there is some chance of misclassification. Gail et al. *(8)* studied designs with 5893 families, chosen to achieve ±5% precision on $\phi_1 = 0.9$, with $\phi_0 = 0.1$ and $q = 0.01$. They defined sensitivity as the chance that a diseased relative will be described as diseased by the proband and specificity as the chance that a nondiseased relative will be described as nondiseased. A sensitivity of 0.9 induced a downward bias of about 10% in $\hat{\phi}_0$ and $\hat{q}$ and a downward bias of 3% in $\hat{\phi}_1$. A specificity of 0.9 induced large upward biases; average estimates of $\hat{\phi}_0$, and $\hat{q}$ (with standard errors) in 1000 simulations were 0.945 (0.016), 0.184 (0.004), and 0.017 (0.002), respectively.

Violation of the conditional independence assumption A3. Gail et al. *(8)* examined a logistic model that included a normally distributed familial effect, b, with mean zero and variance τ^2. The random familial effect is drawn independently for each family, and each member of a family shares the same random effect. A member of a family with carrier status g and familial random effect b has a probability of disease

$$P(Y = 1 \mid g, b) = \{1 + \exp(-\mu_g - b)\}^{-1}, \qquad [3]$$

where μ_1 and μ_0 correspond to $g = 1$ and 0 and are chosen so that the marginal probabilities satisfy $P(Y = 1 \mid g = 1) = 0.9$ and $P(Y = 1 \mid g = 0) = 0.1$. For example, for $\tau^2 = 4.0$, $\mu_1 = 3.4095$, and $\mu_0 = -3.4095$.

The random familial effect induces residual correlations among family members' phenotypes, conditional on their genotypes. For example, with $\tau^2 = 4.0$, the correlation between the phenotypes of two family members, each with genotype $g = 1$, is 0.287. To determine the effects of such familial effects, Gail et al. *(8)* simulated 1000 studies, each based on 2178 families, 10% of which had case probands. Mean estimates of ϕ_1, ϕ_0, and q (with standard errors) from the standard model without random effects were 0.948 (0.026), 0.087 (0.005), and 0.015 (0.002), respectively. Thus, the random familial effect leads to overestimation of $P(Y = 1 \mid g = 1)$ and q and underestimation of $P(y = 1 \mid g = 0)$. It is as if estimates of parameters of the genetic model are exaggerated to accommodate the component of familial aggregation not imparted by the gene under study.

We repeated these analyses using parameter estimates based on the data of Struewing et al. *(2)*; namely, $\phi_1 = 0.56$, $\phi_0 = 0.13$, and $q = 0.0116$. Setting $\tau^2 = 4.0$ in the previous random effects model, we required $\mu_1 = 0.3980$ and $\mu_0 = -2.9948$ to preserve the marginal probabilities $P(Y = 1 \mid g = 1) = 0.56$ and $P(Y = 1 \mid g = 0) = 0.13$. Mean estimates of ϕ_1, ϕ_0, and q (with standard errors) from the standard model without random effects were 0.668 (0.046), 0.113 (0.003), and 0.014 (0.001), respectively. Thus, ignoring the random effect increased the estimated penetrance for carriers and the allele frequency, while it decreased the estimated penetrance for noncarriers. These results resemble those in the previous example.

An independently segregating mutant allele, C, could also induce residual familial correlation given genotypes for A and a. Gail et al. *(8)* found that, even if C had high penetrance, it induced little bias in estimates of $P(Y = 1 \mid g = 1)$, $P(Y = 1 \mid g = 0)$, and q from a naive model based only on alleles A and a, provided C was uncommon. In contrast, simple segregation analysis of these data would have yielded results that combined the effects of A and C and estimated a combined frequency of both A and C. Because the kin-cohort analysis is based on genotyping the proband for A and a and because a rare allele C induces relatively little residual correlation, serious bias is avoided.

Violation of HWE. Gail et al. *(5,8)* studied a population consisting of two noncommunicating strata within each of which mating was random and HWE held but between which there was no mating. Therefore, HWE does not hold in the overall population. The allele frequencies within strata were chosen to preserve the carrier frequency in the entire population. Gail et al. *(5,8)* found little evidence for bias in estimates of ϕ_1, ϕ_0, or q in the presence of such violations of HWE, provided allele frequencies were small, because the joint distribution of Y_1, Y_0, g_1, and g_0 was little affected by such stratification.

Validity of asymptotics. As indicated by the examples in the "Sample Size" section, large samples are needed to achieve good precision for estimates of ϕ_1 when allele frequencies are rare. Fig. 1, taken from Gail et al. *(8)*, shows a histogram of values of $\hat{\phi}_1$ based on 5000 simulations, each with 589 case probands and 5304 control probands and with $m = 2$ relatives (daughters) in each family. These numbers of probands were chosen to achieve ±5% precision on $\hat{\phi}_1$. The distribution of $\hat{\phi}_1$ is symmetric (Fig. 1), and the average estimates (with standard errors) of $\phi_1 = 0.90$, $\phi_0 = 0.10$, and $q = 0.01$ were 0.900 (0.026), 0.100 (0.003), and 0.010 (0.001), respectively. The coverage of a nominal 95% Wald confidence interval for $\phi_1$was 0.932, and that of a profile likelihood-based confidence interval was 0.945.

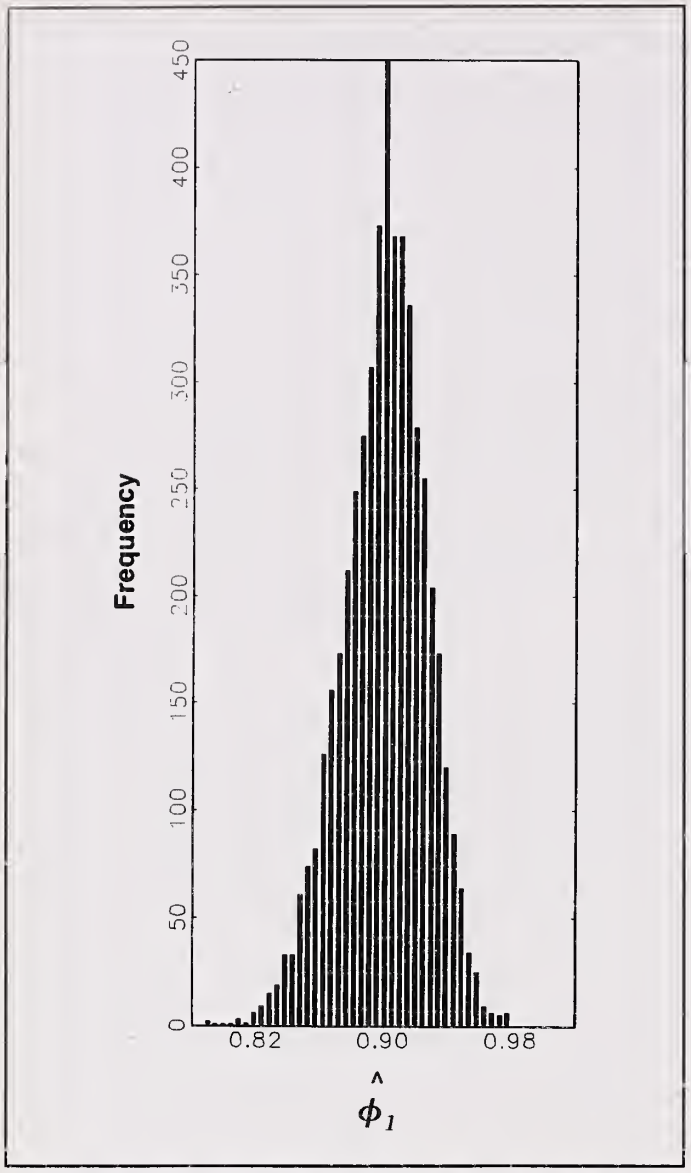


Fig. 1. Histogram of $\hat{\phi}_1$ for samples with 5893 families, 10% of which have case probands. The corresponding parameter values are $\phi_1 = 0.9$, $\phi_0 = 0.1$, and $q = 0.01$.

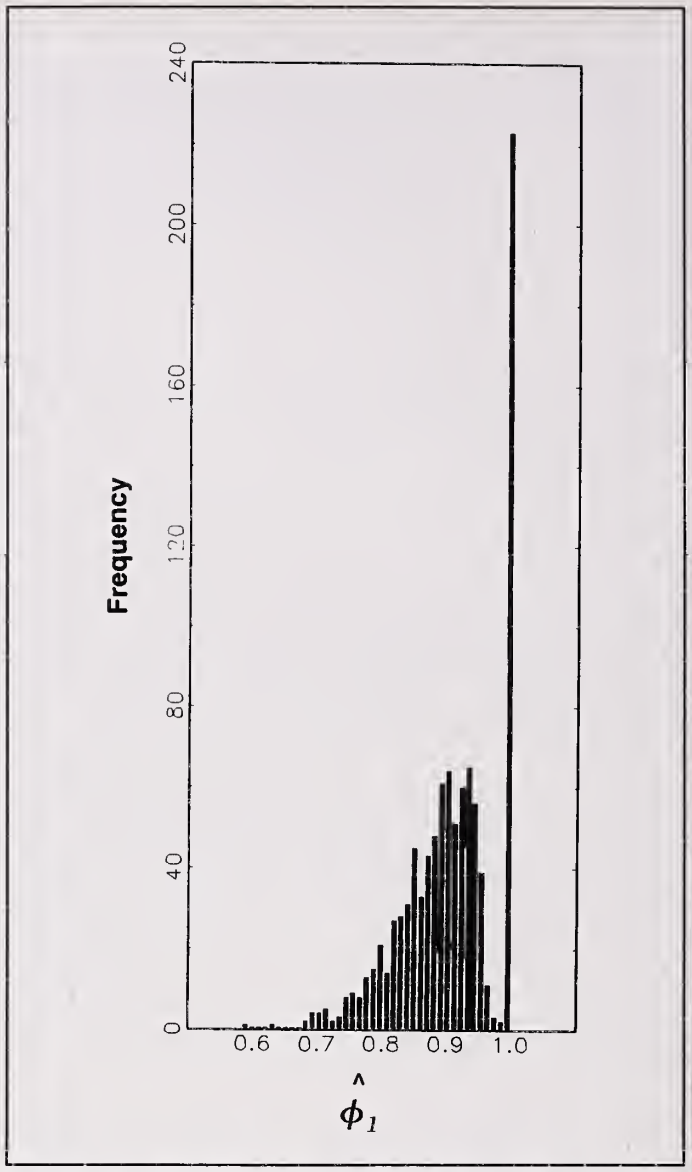


Fig. 2. Histogram of $\hat{\phi}_1$ for samples with 737 families, 10% of which have case probands. The corresponding parameter values are $\phi_1 = 0.9$, $\phi_0 = 0.1$, and $q = 0.01$.

By contrast, a smaller study with 74 case probands and 663 control probands often yielded estimates $\hat{\phi}_1$ on the boundary $\hat{\phi}_1 = 1.0$ (Fig. 2). The small sample distribution of $\hat{\phi}_1$ was distinctly non-normal, and the Wald confidence interval had coverage 0.773, based on 1000 simulations. It is encouraging that the profile likelihood-based confidence interval for ϕ_1 had coverage 0.942.

Survival data. We investigated whether survival data would be subject to similar selection bias and bias from ignoring violations of the conditional independence assumption A3. We assumed that the time to cancer onset among carriers followed a Weibull distribution, $P(T \geq t) = \exp\{-(\lambda t)^{\alpha}\}$, where $\lambda = 0.013024$ and $\alpha = 2.133435$ were chosen to match the cumulative risk of developing breast cancer to ages 50 and 70 years found by Struewing et al. *(2)*, namely, 0.33 and 0.56. Likewise, a Weibull model with $\lambda = 0.007845$ and $\alpha = 3.289313$ matches the cumulative risk to ages 50 and 70 years of noncarriers found by Struewing et al. *(2)*, namely, 0.045 and 0.13. The allele frequency was $q = 0.0116$. The simulated potential ages of the sister proband, sister, and mother were assumed to be multivariate normal with means and covariances chosen to match data in Struewing et al. *(2)*, and censoring from competing mortality and from death following breast cancer onset was taken into account as described by Gail et al. *(5)*.

We assessed 1 : 2 : 4 selection bias (*see above*) from 10 simulations, each including 10 000 families, of which 25% had case probands. The mean cumulative risk to age 70 years in carriers (with standard error) was 0.705 (0.022), the mean cumulative risk in noncarriers was 0.209 (0.003), and the mean estimated allele frequency was 0.018 (0.002). Thus, as for dichotomous outcomes, 1 : 2 : 4 selection bias led to overestimation of the cumulative risk in carriers and noncarriers and of the allele frequency.

We studied violation of the conditional independence assumption by multiplying all hazards within a given family by a frailty factor, γ, that has a chi-squared distribution with 1 *df*.

Thus, γ has mean 1 and variance 2. Weibull parameters were adjusted to preserve marginal probabilities in the presence of this frailty, of $P(T \leq 50 \mid g = 1) = 0.33$, $P(T \leq 70 \mid g = 1) = 0.56$, $P(T \leq 50 \mid g = 0) = 0.045$, and $P(T \leq 70 \mid g = 0) = 0.13$. We conducted 100 simulations each based on 5000 families, half of which had carrier probands. Using the naive survival model without frailty, we observed average cumulative risks (with standard errors) to age 70 of 0.681 (0.060) for carriers and 0.12 (0.003) for noncarriers and an average estimated allele frequency of 0.014 (0.001). Thus, as for dichotomous data, ignoring residual familial correlation led to overestimation of the penetrance for carriers, underestimation of the penetrance for noncarriers, and overestimation of the allele frequency.

Discussion

The kin-cohort design has several practical advantages. In favorable circumstances, this design can be implemented quickly, compared with cohort or population-based case–control designs because information on relatives can be obtained from a single interview of the proband and because one is relying on the previous history of disease in the proband and relatives to estimate penetrance. Moreover, information on several diseases can be obtained from a single proband interview, and several mutations can be studied by appropriate genotyping of the proband. Gail et al. *(5)* found that the kin-cohort design usually requires slightly smaller sample sizes than cohort or population-based case–control designs to achieve the same precision on estimates of the penetrance among mutation carriers, and the reductions in sample size can be considerable if some relatives of the proband can also be genotyped.

Despite the potential practical advantages of the kin-cohort design, it is susceptible to certain biases that affect other forms of segregation analysis. In particular, penetrances and allele frequencies can be seriously overestimated if the probability that a potential proband participates increases with the number of affected relatives of the proband. Two-stage designs that first determine the number of affected relatives before attempting to recruit a proband afford the possibility of bias correction *(9)*. Standard population-based case–control designs and cohort designs are relatively impervious to this type of ascertainment bias *(5)*.

Imperfect specificity in determining the phenotypes of relatives also leads to overestimates of penetrance and allele frequencies. This finding highlights the desirability of applying the kin-cohort design only in populations that can provide good family history data or in special settings with registry data that can provide the needed information on phenotypes of family members *(1)*. Such a study of a founder mutation in BRCA2 in Iceland, where there is a cancer registry and recorded links to family members, yielded estimates of cumulative breast cancer risk to age 70 years of 37% *(11)*.

Residual familial correlations induced by factors such as common dietary habits can lead to an exaggerated estimate of the effects of the gene under study and, in particular, to overestimates of $P(Y = 1 \mid g = 1)$ and q and underestimates of $P(Y = 1 \mid g = 0)$. We found similar results for survival data. On the other hand, another rare gene that segregates independently has little effect on estimates of these parameters for the gene under study in the kin-cohort design.

Violations of HWE have little effect on estimates of penetrance.

Very large samples may be needed to ensure the validity of standard asymptotic approximations. In some examples, however, likelihood-based confidence intervals for ϕ_1 had near-nominal coverage even when the distribution of $\hat{\phi}_1$ was evidently non-normal. This problem can be alleviated by studying special populations with higher allele frequencies because the number of families needed to attain precise estimates and good performance of asymptotic approximations is reduced in such populations.

In many ways, the study of the penetrance of mutations of BRCA1 and BRCA2 in the Washington Ashkenazi study by Struewing et al. *(2)* was an ideal application for the kin-cohort design. Indeed, Wacholder et al. *(1)* developed the design with this study in mind. The prevalence of mutant alleles of about $q = 0.01$ in this population reduced the required sample sizes, compared with a study in the general population with allele frequency 0.003, say. Most of the probands were well educated and could provide good information on the breast cancer status of first-degree relatives. Although there was some evidence that the probands had more affected relatives than in the general population of Ashkenazi women *(2)*, the effect of such an ascertainment bias would be to overestimate penetrance. The striking result of this study, however, was that the estimate of penetrance was lower than that obtained from highly affected pedigrees *(3)*, a result that cannot be explained by such a bias.

An important concern in studies of gene penetrance is the need to respect the autonomy of study participants and to protect them from harm that might result from the improper use or release of confidential genetic information. One issue is whether the genotyped participants will be offered genotype information. This should only be done in studies that provide counseling to determine the wishes of the participants and to convey genetic information appropriately. The institutional review board overseeing the study by Struewing et al. *(2)* determined that genotypes should not be revealed to participants and required that all data linking genotypes with individuals be eliminated after the main analysis. Kin-cohort designs that seek additional interview data or samples from relatives selectively on the basis of the genotype of the proband may inadvertently convey unwanted information on the risk of carrying an adverse mutation, either to the proband or to the relatives. Special care would be needed, therefore, in obtaining informed consent for such designs.

One problem that arose in the context of the study by Struewing et al. *(2)* was that more than one volunteer proband would sometimes come from the same family. In this case, it is not evident how to modify the likelihood *(1)*. Although standard methods can be used to obtain representative samples of individual probands, it is not clear how to sample families without reference to a proband. Indeed, although the set of relatives of a sampled proband is well defined, it is not clear how to define a family without reference to a proband. Thus, there are some conceptual issues that require further clarification for population-based inference when one goes beyond the framework of representative sampling of probands.

In principle, the kin-cohort design has other potential applications *(1)*. One can introduce covariates into the penetrance function and use regression methods to estimate genetic effects, the effects of measured covariates, and their interactions. A potential weakness of this strategy may be in obtaining accurate data on covariates in relatives. The kin-cohort design can also be used to study the main effects of two independently segregating

genes, as well as their interactions. Particularly large samples might be required to test for an interaction if only the proband is genotyped, however, because the genotype of the proband gives only partial information on the genotype of the relatives. Other potential applications include evaluation of the effects of a particular gene on other characteristics of the relative, such as their survival following cancer onset *(12)*. It would seem that a more powerful approach for such studies would be to genotype the relatives rather than to rely on less direct information from the proband's genotype, but ethical or cost constraints may prevent genotyping relatives, and some of them may have died.

There is scope for further methodologic work for survival models. Wacholder et al. *(1)* took a nonparametric approach. They recognized that the survival curve (1 – the cumulative risk function) of a relative of a proband who is a carrier is a mixture (approximately 50 : 50) of survival curves for carriers and noncarriers, whereas the survival curve for a relative of a proband who is a noncarrier is another mixture (approximately 0 : 100) of survival curves for carriers and noncarriers. Using Kaplan–Meier estimates of the observable mixed survival distributions for relatives of carrier probands and for relatives of noncarrier probands, they solved two linear equations for the underlying survival distributions for carriers and noncarriers. Struewing et al. *(2)* were not explicit about how allele frequencies were estimated, and these allele frequencies are needed to define the mixing coefficients. Nonetheless, the solutions are insensitive to the exact allele frequencies for rare alleles. Although the procedures used in the study by Struewing et al. are consistent, provided estimates of allele frequencies are consistent, in small samples they can lead to nonmonotonic estimates of the survival curves for carriers and noncarriers.

In principle, parametric approaches, such as those used by Gail et al. *(5)* for a three-parameter improper Weibull model for carriers and for noncarriers, can be used for inference in a manner similar to that described in the "Survival data" section. In moving toward a nonparametric solution, one could consider piecewise exponential models instead, as used by Claus et al. *(6)*. These approaches have the advantage that estimated survival curves will be monotonic, provided parameters are fit subject to their natural constraints. As the number of parameters increases, however, full likelihood methods may become unstable and may not converge, and there is no guarantee of consistency if the number of parameters increases at the same rate as the sample size. Thus, further work is needed to develop survival estimates that are weakly parametric or nonparametric and to allow for potential residual familial eorrelation.

References

(1) Wacholder S, Hartge P, Struewing JP, Pee D, McAdams M, Brody L, et al. The kin-cohort study for estimating penetrance. Am J Epidemiol 1998;148: 623–30.

(2) Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, et al. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 1997;336:1401–8.

(3) Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Am J Hum Genet 1995;56:265–71.

(4) Whittemore AS, Gong G, Itnyre J. Prevalence and contribution of BRCA1 mutations in breast cancer and ovarian cancer: results from three U.S. population-based case–control studies of ovarian cancer. Am J Hum Genet 1997;60:496–504.

(5) Gail MH, Pee D, Benichou J, Carroll R. Designing studies to estimate the penetrance of an identified autosomal dominant mutation: cohort, case–control and genotyped-proband designs. Genet Epidemiol 1999;16:15–39.

(6) Claus EB, Risch NJ, Thompson WD. Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. Am J Hum Genet 1991;48:232–42.

(7) Thomas DC. Design of gene characterization studies: an overview. Monogr Natl Cancer Inst 1999;26:17–23.

(8) Gail MH, Pee D, Carroll R. Properties of likelihood methods for estimating the penetrance of an autosomal dominant mutation in kin-cohort studies. J Statist Planning Inference. In press 1999.

(9) Siegmund KD, Whittemore AS, Thomas DC. Multistage sampling for disease family registries. Monogr Natl Cancer Inst 1999;26:43–8.

(10) Cannings C, Thompson EA. Ascertainment in the sequential sampling of pedigrees. Clin Genet 1977:12:208–12.

(11) Thorlascius S, Struewing JP, Hartge P, Olafsdottir GH, Sigvaldason H, Tryggvadottir L, et al. Population-based study of risk of breast cancer in carriers of BRCA2 mutation. Lancet 1998;352:1337–9.

(12) Lee JS, Struewing JP, Wacholder S, McAdams M, Pee D, Brody LC, et al. Survival after breast cancer in Ashkenazi Jewish BRCA1/2 mutation carriers. J Natl Cancer Inst 1999;91:259–63.

Note

Supported by Public Health Service grant CA57030 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services (R. Carroll).

III. INTEGRATION PANEL

Study Design in Genetic Epidemiology: Theoretical and Practical Considerations

Alice S. Whittemore, Lorene M. Nelson

Recent advances in molecular genetics have created new opportunities and challenges for genetic epidemiologists. Here we review some of the issues that arise when designing a study involving the genetic epidemiology of chronic diseases of late onset, such as cancer. We discuss two considerations that influence the choice of design. The first consideration is the study's goals. We describe the goals of identifying new susceptibility genes for a disease, of estimating important characteristics of known genes, and of learning how to prevent the disease in the genetically susceptible. We indicate how these goals affect the choice of design and present some guidelines for choosing designs that effectively address them. The second consideration is the set of practical constraints to successfully conducting the research. These contraints include problems of potential selection bias, reduced response rates, problems particular to family registries, problems particular to the cultures of various ethnic groups, and ethical issues. We indicate how these constraints affect the choice of design and discuss ways to deal with them. [Monogr Natl Cancer Inst 1999;26:61–9]

Technologic developments in molecular genetics have created new opportunities and challenges for genetic epidemiologists. Within the past three decades, it has become feasible to localize and sequence a disease-susceptibility gene before its function is known. Such gene identification is accomplished using one or more of several strategies, collectively known as *positional cloning.* These strategies include linkage and association studies and, for site-specific cancers, studies of allelic loss and differential gene expression in cancer cells compared with normal cells in the same individual. Positional cloning of genes using one or more of these techniques leads immediately to basic biologic questions. For example, we need to know the function of the protein coded by the gene and how polymorphisms of the gene lead to functional variants in its protein that alter disease risk. Answers to these questions, if available, can suggest important gene–environment or gene–gene interactions and can motivate preventive strategies.

An alternative approach to positional cloning for gene identification is evaluation of *candidate genes,* i.e., those coding for proteins whose functions are thought to be involved in the disease process. By definition, the function of a candidate gene is at least partially understood. Successful gene identification, whether by positional cloning or by implication of candidate genes, introduces several new areas of inquiry for the genetic epidemiologist. For example, a question of practical clinical importance concerns the age-specific disease risks associated with the various genotypes. A question with less immediate clinical consequences, but one that is important for public health planning and resource allocation, concerns the population frequencies of genotypes associated with increased risk. A related issue is estimation of that fraction of the disease burden due to variant genotypes.

When several genes have been identified, we need to know how these genes interact to affect disease risk. Perhaps the most important issue, from a public health perspective, concerns how lifestyle characteristics modify risk in carriers of high-risk genotypes. Specifically, what can such carriers do to reduce their risk? Why do some carriers remain disease free well into old age? Studies aimed at investigating these areas of inquiry, which we shall call *gene characterization* studies, present challenging issues for genetic epidemiologists and will continue to do so well into the 21st century.

This paper considers some of the issues involved in designing studies to identify and characterize disease genes. We focus on two considerations: 1) the goals and objectives of the study and 2) the practical constraints to conducting the research. We indicate how these two considerations affect the choice of design, and we summarize some guidelines for choosing designs that effectively address the study objectives while circumventing the limiting constraints.

Effect of Scientific Goals on Design

The goals of a study typically are to identify new genes that predispose to the disease of interest or, if a predisposing gene is known, to estimate the risks associated with specific alleles and the prevalences of those risky alleles and to determine the way that other genes or lifestyle characteristics modify those risks. An additional important goal is to evaluate disease-prevention strategies in the genetically susceptible. Currently, for example, uncertainty exists over the advisability of prophylactic mastectomy or oophorectomy or both in female carriers of the genes BRCA1 and BRCA2, largely because of the lack of data comparing the benefits of such surgery (cancer risk reduction) with its major physical and psychologic costs. However, a retrospective cohort study of 639 women with a family history of breast cancer who underwent bilateral prophylactic mastectomy at the Mayo Clinic between 1960 and 1993 and who were followed for subsequent breast cancer incidence showed a 90% reduction of incidence compared with that predicted in the absence of surgery *(1).*

Identifying New Genes

Studies aimed at gene identification fall into several categories. These categories include (*a*) genome-wide scans using

Affiliation of authors: Department of Health Research and Policy, Stanford University School of Medicine, CA.

Correspondence to: Alice S. Whittemore, Ph.D., Department of Health Research and Policy, HRP Redwood Bldg., Rm. T204, Stanford University School of Medicine, Stanford CA 94305-5405 (e-mail: alicesw@leland.stanford.edu).

See "Notes" following "References."

© Oxford University Press

parametric and nonparametric linkage analyses, (*b*) association studies using cases and unrelated controls, (*c*) genome-wide scans using family-based association studies, (*d*) evaluation of candidate genes, and, for cancer, (*e*) allelic loss and differential gene expression studies based on comparison of cancer cells and normal cells in the same individual. Choosing among these designs requires that we consider several factors, including the anticipated frequencies and penetrances of alleles at the disease locus, the numbers of alleles and frequencies of alleles at nearby markers, and the disease phenotype of interest (e.g., multiple cancers or intermediate biomarkers). Genome-wide linkage and association scans exploit the recent availability of a large set of DNA variants, called *markers,* each having a known location on a particular chromosome. A marker might be a gene, a sequence of bases that varies from one chromosome to another in a detectable way, or a single nucleotide polymorphism (SNP). The chromosomes of family members with and without the disease are labeled with respect to their observed alleles at each of the markers.

Genome-Wide Scans Based on Linkage

Linkage or allele-sharing studies exploit the fact that markers having some alleles shared by the diseased family members and other alleles shared by the nondiseased members are likely to be proximal to a disease-susceptibility gene. When such disease-specific sharing of the marker alleles is unlikely to be a result of chance, the chromosomal region around the marker(s) is searched intensively in an attempt to identify a critical subregion containing the disease gene and, indeed, the gene itself. Association studies, by contrast, search for the correlation between disease and specific marker alleles in much the same way as classic epidemiologic studies search for the correlation between disease and personal attributes.

Parametric linkage analysis requires a genetic model for both the disease phenotype and the marker genotypes. This analysis includes specification of the locations of the markers, with an unknown parameter representing the location of the putative disease gene. This parameter, which is of primary interest, is estimated from data on observed disease phenotypes and marker genotypes in families, by the method of maximum likelihood. In contrast, nonparametric or "model-free" linkage analysis, also called allele-sharing analysis, uses the family data on disease phenotypes and marker genotypes to evaluate departures from the null hypothesis that the marker genotypes are not near a disease gene. Whether parametric or nonparametric, linkage analysis is concerned not with association between disease phenotype and a particular marker allele, but rather with patterns of marker allele sharing within diseased and nondiseased family members. Suppose, for example, families are typed at a marker with alleles A and B that is close to a disease gene. Then allele A may segregate with disease in some of the families, while allele B does so in others; the clues come from *patterns* of allele segregation with disease rather than association of specific alleles with disease. A more detailed discussion of linkage studies and issues in their design can be found in the paper by Schaid et al. *(2)* in this monograph.

Association Studies With Case Patients and Unrelated Control Subjects

An alternative to linkage for gene identification is a set of statistical methods that we shall call association analyses. The simplest is the classic case–control design that compares marker genotypes of diseased case patients with those of unrelated disease-free control subjects. With current technology, the case–unrelated control design is used to examine a small number of candidate gene polymorphisms as they relate to risk of the disease under study. This type of study design is appropriate for the study of relatively common alleles that confer a modest or small effect on disease risk. In the near future, technology will make it possible to conduct whole-genome association studies that compare allele frequencies across the genome between case patients and unrelated control subjects. Once the catalog of human DNA sequence variation is complete, it will be possible to directly test the common functional variants that occur in coding regions (expecting that two or three variants each with a frequency of 10% or greater will occur in a given coding region) and to compare these allele frequencies between case patients and unrelated control subjects *(3).* A second approach will involve the use of SNPs in a dense map that spans both the coding and noncoding regions of the human genome *(3).* This approach will be based on the rationale that affected individuals will be characterized by a particular haplotype in the neighborhood of the susceptibility gene and that comparison with unaffected control subjects will identify specific regions where there is an increased prevalence of a particular set of SNP alleles in affected individuals. The availability of approximately 100 000–200 000 SNPs will allow a dense map for the identification of susceptibility genes; however, the large number of statistical comparisons will require large sample sizes to account for the increased probability of identifying associations that occurred on the basis of chance alone.

Investigators have worried that measures of association obtained from studies that compare case patients and unrelated control subjects may be biased by confounding because of ethnic stratification of the population. Suppose, for instance, that the locus under test is not near a disease locus, but that the population consists of two ethnic groups, with the first group having higher prevalence of both the disease and a specific allele than the second group. A random sample of eases would contain a higher fraction of the first ethnic group than would either the general population or a sample of controls and, thus, a higher total count of the allele. In this situation, the test statistic will lead to rejection of the null hypothesis more often than it should. Thus, failure to account for such *ethnic stratification,* either by a matched design or by a stratified analysis, could lead to invalid conclusions. Adjusting for population stratification requires collecting detailed race or ancestry information; however, this approach may leave residual confounding because many genetic traits vary in frequency within apparently homogenous ethnic groups. Race or ethnicity is also an incomplete surrogate for genetic makeup because there may be cultural and geographic influences on risk within subgroups of a given racial or ethnic group. Even within an apparently homogenous ethnic group, there may be different allele frequencies at many loci because of the effects of different geographic locations or migration patterns, and both of these factors may affect ethnic admixture within the apparently homogenous group. Thus, it is advisable to compare cases and controls within narrowly defined ethnic groups. Such comparisons may be difficult to achieve in the United States because of the increasing proportion of individuals with mixed race or ancestry. It is worth noting that the potential bias of concern here is a classic example of confounding (by

ethnic origin), a problem that has been well studied by epidemiologists. In particular, it has been shown *(4)* that the confounding factor (ethnic ancestry) can produce substantial bias only if it is both strongly associated with disease risk and strongly correlated with genotypes at the disease-associated locus. Moreover, large case–control differences in prevalences of suspect genotypes at the locus are less likely than small differences to be due to such bias. These issues are discussed in more detail by Caporaso et al. *(5)* in this monograph.

Genome Scans Using Family-Based Association Analyses

The problem of confounding by ethnic stratification in case–control studies that use unrelated control subjects has prompted interest in family-based designs, such as comparison of genotypes of affected and unaffected siblings or comparison of genotypes of affected offspring with those of their parents *(6)*. These designs automatically match "controls" to cases on ethnic ancestry. For example, much has been written about the transmission disequilibrium test that compares the marker alleles transmitted from heterozygous parents to affected offspring with those not transmitted [*see* Schaid *(7)* for a review and commentary]. In this monograph, Gauderman et al. *(6)* note that family-based designs are generally less efficient than designs based on unrelated control subjects. Research is needed to investigate the trade-offs in bias and efficiency involved in the various design choices for association studies.

An important question in searching for new genes concerns the relative merits of linkage and association analyses for genome scans. Risch and Merikangas *(8)* argue that, whereas the former have more power for genes with rare, highly penetrant disease-susceptibility alleles such as are often observed among families with autosomal dominant inheritance, association studies are preferred for genes having common alleles with low penetrance. Their conclusions are based on the assumption that the disease-causing and marker polymorphisms are identical. This assumption has been questioned, and consequences of its violation have been investigated *(9–11)*. In situations in which the disease and the marker polymorphisms are not identical even within the same gene, if the extent of disequilibrium between them is not perfect and if the frequencies of alleles at the two loci differ substantially, association analyses can be less powerful than linkage analyses even for common, low-penetrance polymorphisms. Thus, the extent of disequilibrium between disease-causing and marker loci is a critical determinant of the power of association studies when genotypes at the disease locus have moderate to low penetrance. Another issue that must be considered when contemplating a genome-wide association scan is the likely etiologic basis of the disease of interest. Complex diseases, in which environmental and behavioral factors combine with genetic factors to cause increased risk, must be investigated in more complex ways than monogenic or oligogenic diseases.

Evaluation of Candidate Genes

To avoid inflated type-I error probabilities, the genome scans described above must accommodate the multiple testing involved in evaluating the 300 to 100 000 or more markers evaluated in a scan. In contrast, investigations of candidate genes, based on an *a priori* biologic hypothesis that a particular gene or family of genes is involved in the disease, avoid this problem. Examples of candidate genes include those involved in estrogen metabolism in relation to breast cancer; in the synthesis, metabolism, and response to testosterone in relation to prostate cancer; and in the metabolism of heterocyclic amines from well-cooked meat in relation to colon cancer, as described by Le Marchand *(12)* in this monograph. Candidate genes can be examined with the use of the analytic methods described above for association studies. This approach suffers from a drawback similar to that noted above for association-based scans, namely, lack of knowledge of the exact polymorphism that is causally related to disease risk. Instead, investigators identify polymorphisms having variation in the general population and then examine them for association with disease. The absence of tight linkage disequilibrium between causal and measured alleles can induce problems similar to those of misclassification error, namely, power loss and attenuation bias in relative risk estimates.

Some issues in gene identification apply to both genome scans and evaluation of candidate genes. For example, there is a need for designs and analytic methods that accommodate the late and variable ages at onset of most chronic diseases. Clearly, a man who dies in his 40s from a myocardial infarction should not be considered unaffected with respect to diseases of late onset, such as prostate cancer. Analyses that account for age, however crudely, should be used when investigating the etiologic basis of diseases of late onset. A related issue, discussed in this monograph by Schaid et al. *(2)*, concerns the need to adjust analyses for personal covariates, such as genotypes at other known predisposing genes, or lifestyle characteristics, such as diet or tobacco consumption. The issue of which covariates to include has implications for data collection procedures. Also, there is a need to assess effect modification of lifestyle or environmental influences with the presence or absence of certain genotypes.

Allelic Loss and Differential Gene Expression Studies

In contrast to linkage and association studies that evaluate genetic characteristics in diseased and disease-free individuals, allelic loss and differential gene expression studies identify cancer-susceptibility genes by comparing genetic characteristics of cancer cells and normal cells from the same individual. Allelic loss studies identify genes whose proteins prevent a cell from becoming malignant. These genes, the tumor suppressor genes, operate either by repairing damage to a cell's DNA or, if repair is impossible, by initiating apoptosis. The gene p53 on chromosome 17p is an example of a tumor suppressor gene. A cell that has lost both functioning alleles of p53 (by mutation or deletion) is at high risk of malignant transformation. A cancer composed of the malignant progeny of such a cell provides an ideal testing ground for finding new genes. If the normal cells in the tissue contain two alleles for markers in the region, whereas the cancer cells have lost one or both alleles, there may be a nearby tumor suppressor gene whose loss is responsible for the cancer. A limitation of these allelic loss studies is that the loss in the cancer cells may merely be a consequence of, rather than a cause of, the malignancy.

Like allelic loss studies, differential gene expression studies compare cancer cells with normal cells. However, the objective is not only to identify tumor suppressor genes but also to identify oncogenes whose inappropriate expression is involved in malignant transformation. Genes detectable in the messenger RNA of cancer cells but absent in that of normal cells in the same tissue are good candidates for oncogenes. Comparative genomic hybridization is but one of many techniques recently developed to evaluate gain or loss of genetic material in cancer cells compared

with normal ones. The Cancer Genome Anatomy Project (CGAP) of the National Cancer Institute aims to provide a public database of differential gene expression in cancer cells and in normal cells.

The likelihood of identifying disease genes is greatest for monogenic diseases, such as cystic fibrosis, or for diseases caused by rare, highly penetrant mutations of a gene, such as breast cancer caused by mutations of BRCA1. More problematic are the complex diseases, such as non-insulin-dependent diabetes mellitus (NIDDM) or autism, most likely caused by mutations of multiple low-penetrance genes. For instance, family and twin studies indicate a strong genetic component in NIDDM, and mathematic modeling suggests the interaction of multiple genes. Most studies in NIDDM have relied on the candidate gene approach that has successfully implicated mutations in the glucokinase gene *(13)* and the glucagen receptor gene *(14)*. As positional cloning becomes more sensitive and powerful, it will complement the candidate gene approach to identifying new genes for complex diseases, although it seems unlikely to supplant it. It also may be necessary to take into account interactions, either among genes or between genes and environmental exposures, to successfully detect additional disease-susceptibility genes.

An important question is when to stop searching for new genes for a specific disease. For a rare disease with a clear mendelian inheritance pattern, such as cystic fibrosis, the answer is straightforward. This inherited disorder, affecting one in 2000 Caucasians, is caused by impaired ion transport because of malfunction of a protein called cystic fibrosis transmembrane conductance regulator. When the gene encoding this protein was cloned in 1989, it was discovered that its mutations cause virtually all cases of the disease. In contrast to the cystic fibrosis story, it is not clear when to stop searching for new genes predisposing to chronic diseases with complex genetic etiologies. For example, it is not known how many distinct genes have polymorphisms with a detectable influence on breast cancer risk. Several investigations *(15–17)* describe families with multiple cases of breast cancer that are segregating mutations of neither BRCA1 nor BRCA2. As another example, it is not known whether all familial aggregation of ovarian cancer can be explained by mutations of BRCA1, BRCA2, and chance. New analytic methods are needed to address this issue.

Characterizing Attributes of Known Genes

Once a gene that affects disease risk has been identified, there are several analytic tasks at hand. These include estimation of (*a*) risks associated with specific alleles, (*b*) frequencies of high-risk alleles in various populations, (*c*) gene–gene and gene–environment interactions, and (*d*) the proportion of the disease burden attributable to polymorphisms of the gene.

As the field of genetic epidemiology burgeons and the number of candidate genes increases, it will be increasingly important to give careful consideration to the rationale for investigating a specific gene mutation or allelic variant. Several possible rationales exist: (*a*) The genetic variant has been associated with familial forms of the disease; (*b*) the gene codes for proteins involved in biochemical mechanisms of disease pathogenesis; (*c*) the gene codes for xenobiotic enzymes that are thought to interact with environmental exposures to affect disease risk; and (*d*) the biologic role of the genetic variant in disease etiology is unclear, but the variant is suspected to be disease associated or has been associated with disease risk in previous studies.

Of primary importance is the strength of the rationale for investigating a given mutation or polymorphic variant. Investigations that are based on rationale (*a*) are clearly worthwhile because there is strong biologic plausibility that the genetic variant is causally related to disease risk. One example is the discovery of BRCA1 mutations in breast cancer. Following the initial discovery of the BRCA1 gene, subsequent investigations *(18,19)* sought to estimate the proportion of breast cancer cases that are attributable to BRCA1 mutations by estimating carrier frequencies among familial and sporadic breast cancer cases and among population controls. Rationale (*b*) also provides a strong basis for investigation if a gene codes for a protein that is known to be involved in disease pathogenesis, as is the case for the amyloid-beta precursor protein in Alzheimer's disease *(20)*. Investigations based on rationale (*c*) may be biologically plausible because xenobiotic enzymes involved in the phase I or II metabolism of chemicals may adversely affect the successful detoxification of environmental carcinogens *(21)*. Genetic variants of these enzymes may have deleterious effects by promoting the rapid metabolism of a compound to its toxic metabolite, or it may confer protective effects by successfully metabolizing the compound to a nontoxic form. Examples are the cytochrome (CYP) P450 enzymes responsible for phase I metabolism of environmental chemicals (e.g., debrisoquine hydroxylase [CYP2D6] and risk of lung cancer) *(22)* and enzymes such as glutathione-*S*-transferase (GST) that are involved in phase II metabolism (e.g., GSTM1 null genotypes and risk of bladder cancer) *(23)*. However, the reasoning regarding a causal role for a given xenobiotic enzyme is rarely straightforward because of the existence of multiple enzymatic pathways for a specific compound and the fact that some xenobiotic enzymes can be induced by other environmental exposures.

Rationale (*d*) is the weakest basis for pursuing the study of a putative susceptibility gene and may be responsible for a burgeoning number of inconsistent findings across studies. Several common flaws in study design can contribute to inconsistent results. The first investigations of a putative susceptibility gene may be based on poorly defined or "convenience" controls or may use allelic frequencies for controls that are based on published literature. As a result, findings from early studies can be extremely variable, and resolution requires more definitive studies with the use of appropriate population controls. A related concern is the failure to collect data or adjust for differences in the racial or ethnic distributions of the case and control groups, as discussed in the previous section. Some studies *(24)* have relied on prevalent cases, which does not allow differentiation of the genetic variants that are associated with better survival (overrepresented among prevalent cases) from variants that increase or decrease the risk of developing the disorder. Characteristics of gene polymorphisms may also affect study feasibility. The population frequencies of the allelic variants may be very low, making it difficult to obtain a sample size large enough to obtain precise estimates of allele frequencies, to estimate main effects, and to estimate gene–gene and gene–environment interactions. Careful consideration must also be given to whether the allelic variants lie in the intron (noncoding) or exon (coding) regions of the gene. If a polymorphism is in an exon (or a close-in promoter region), it is strongly plausible that allelic variants could influence gene expression, and studies addressing the known functional effects of such variants should be consulted. If the allelic variants lie in an intron, they are less likely to have a direct

functional effect but may be in linkage disequilibrium with the true disease susceptibility locus.

In planning investigations of candidate genes, investigators may want to consult an Internet-based resource (Human Epidemiology Genome Network, or HuGENet) offered by the Centers for Disease Control and Prevention (CDC) for on-line publishing of systematic reviews of genetic variants and their associations with the risk of specific diseases. HuGENet is a developing CDC resource for on-line publishing of systematic reviews of certain genetic variants and their associations with the risk of a specific disease. The intent of each HuGENet review is to provide comprehensive background information on molecular or genetic techniques for allelic typing, the prevalence of allelic variants in different racial or ethnic populations, population-based disease risk information, evidence for gene–gene effects or gene–environment interaction, and implications for clinical practice and genetic counseling. Although all of the HugeNet reviews will be peer-reviewed and initially published in peer-reviewed journals such as American Journal of Epidemiology, each on-line review will be updated over time as new information accumulates. The website address for this resource is ⟨http://www.cdc.gov/genetics/hugenet⟩.

Risks Associated With Specific Alleles

Estimates of the age-specific and lifetime risks of disease in carriers of specific mutations or polymorphisms are essential for informed clinical management of those individuals with an inherited disease susceptibility. In principle, cohort studies of identified mutation carriers can be used to estimate their age-specific and lifetime risks, but such studies face many challenges. Unbiased estimates of risk require registries with data on a random sample of carriers. Most existing high-risk families recruited for gene identification studies have been selected for the existence of multiple cancers; therefore, their cancer experience tends to overestimate risk in the general population. In addition, because risk is likely to be influenced by both environmental factors and subsequent interventions, observational studies must gather extensive data on baseline risk factors as well as longitudinal measurements on interventions or changing lifestyles. The pros and cons of classic epidemiologic designs for gene characterization are discussed in more detail by Langholz et al. *(25)* in this monograph. In addition, the family-cohort (also called kin-cohort) design, as described by Struewing et al. *(26)* and Wacholder et al. *(27)*, has been used to assess the penetrances of mutations. This design is discussed in more detail in the paper by Gail et al. *(28)* in this monograph.

Frequencies of High-Risk Alleles

Following the identification of a candidate gene and its association with a given disease, the next step is to obtain accurate estimates of the frequency of high-risk alleles in the population. This information is important for estimating the fraction of the disease that is attributable to high-risk alleles (*see* attributable fraction section *below*) and for determining sample size requirements for subsequent association studies. For gene polymorphisms that are low in frequency (i.e., <10%), large population samples will be needed to obtain precise estimates of the frequencies of high-risk alleles. Furthermore, allele frequencies often vary according to racial or ethnic ancestry, making it very important to collect detailed ancestry information from study subjects so that race-specific allele frequencies can be estimated. In association studies, large sample sizes will be required to estimate the main effects of low-prevalence polymorphisms and to examine gene–gene or gene–environment interaction.

Gene–Gene and Gene–Environment Interactions

When several genes have been implicated in a disease and when the frequencies of double mutant carriers are not negligible, we need to know how the genes interact. At present, for example, little is known about possible interactions between the glucokinase and glucagen receptor genes in NIDDM. And, although an interaction in ovarian cancer has been reported between the genes BRCA1 and a polymorphism close to the H-RAS oncogene *(29)*, little is known about the nature of this interaction, if indeed it is found in other independent data. A more complex gene–gene–environment interaction has been observed in research on bladder cancer. Taylor et al. *(30)* reported evidence of a statistically significant gene–gene–smoking interaction in which the combination with the genotype for slow acetylator status of the *N*-acetyl-transferase (NAT2) enzyme and the presence of one or more NAT1*10 alleles increased the risk for bladder cancer among smokers but not among nonsmokers.

The possibility that modifiable lifestyle characteristics influence gene expression provides hope for the development of preventive strategies. For example, the available data suggest that oral contraceptive use for several years before menopause reduces risk of the type of malignancy that occurs in patients with ovarian cancer without a family history of the disease. The level of protection among carriers of mutations of the BRCA genes is an important issue. The first available data on this issue suggest that oral contraceptive use also reduces the risk of ovarian cancer in carriers of BRCA mutations *(31)*, although its effect on their risk for breast cancer is more problematic *(32)*. Another example is provided by the report in this monograph by Le Marchand *(12)* on gene–diet interaction in Japanese Americans.

Investigators are beginning to evaluate the relative efficiencies of the various study designs for the detection of interaction among known genes and between known genes and endogenous and exogenous characteristics. Reviews of this issue can be found in three articles *(33–35)* and in papers in this monograph *(6,36)*. It is well known that large sample sizes are needed for adequate power to detect interactions among environmental risk factors, and this size requirement is also true for gene–gene and gene–environment interaction.

Attributable Risks

The public health relevance of a given gene polymorphism is addressed by estimating the proportion of diseased individuals in the population that could be prevented if the high-risk alleles were absent (known as attributable fraction, etiologic fraction, or population attributable risk percent). Accurate estimation of the population frequency of the high-risk genotype is important because attributable fraction is a function of the frequency of the high-risk genotype in the population and the penetrances of the high- and low-risk genotypes. Also, because all of these factors may vary with age, it is important to compute age-specific estimates of attributable fraction. Attributable fractions can also be used to estimate the proportion of disease that is a result of the interaction of a genetic variant and an environmental exposure. Because genotypes are not usually modifiable, the prevention of

disease will depend on interventions that target environmental factors that interact with genetic susceptibility to influence the risk of disease. Study designs for assessing interventions among carriers of high-risk genotypes are the topic of the next section.

Evaluating Disease Prevention Strategies in the Genetically Susceptible

As more genes with predisposing alleles are identified and as people become increasingly aware of genetic developments and interested in knowing about their own genes, there is need to offer them options for preventing the diseases to which they are particularly susceptible by inheritance. To achieve this goal, we must understand the functions of proteins encoded by the normal alleles of the disease-related genes, so that we can devise ways to compensate for these proteins when they are dysfunctional and to eliminate as much of the guesswork as possible in devising and evaluating new interventions. The task of evaluating new interventions presents many design challenges, which are beyond the scope of this paper. The following is a brief discussion of issues in the use of two designs for evaluation: observational cohort studies and randomized trials.

Observational Cohort Studies

It is unlikely that observational cohort studies of carriers can be used to measure the relative benefits of two interventions. Bias can arise in an observational study of interventions when patients at higher risk select more aggressive interventions, as might be the case, for example, if patients with BRCA1 mutations and with a high incidence of familial breast cancer tend to select prophylactic mastectomy more often. Bias also can arise if large treatment centers that see a disproportionate share of high-risk patients offer only a single intervention option to identified carriers.

Randomized Trials

The feasibility of randomized trials to compare interventions depends strongly on the disease and on the estimated risk of study participants. The problem of conducting a randomized trial comparing interventions in breast cancer illustrates some of the issues. It seems unlikely that a randomized trial comparing prophylactic mastectomy with a nonsurgical intervention could be completed successfully. Indeed, any such trial will need two active arms, and it cannot contain a no-treatment control. Presently, there is not widespread agreement on a potentially effective nonsurgical intervention for young, high-risk women. There are, however, more options for randomized trials of interventions in carriers of the genes for hereditary nonpolyposis colon cancer. There is a strong rationale for nonsurgical interventions such as nonsteroidal anti-inflammatory drugs (e.g., aspirin) and dietary supplements with calcium or folic acid. Moreover, colonoscopy followed by sampling of the mucosal wall of the colon can be used for the periodic evaluation of study subjects *(37)*.

The recently completed tamoxifen-versus-placebo and the ongoing raloxifene-versus-tamoxifen breast cancer prevention trials select women at high risk of breast cancer and thus contain higher prevalences of BRCA1 and BRCA2 mutation carriers than in the general population. Still, it is not clear that these trials contain sufficient numbers of carriers to evaluate the efficacy of these interventions among carriers.

Practical Constraints

The practical constraints faced by the genetic epidemiologist include those that impede classic epidemiologic research on the etiology of environmental exposures in chronic diseases. These constraints include the need to minimize selection bias and response bias, to ensure accurate and complete data collection and appropriate statistical analysis, and to interact responsibly with all study participants. However, the genetic epidemiologist also faces constraints special to genetic research. These constraints include problems particular to family studies and family registries as well as problems particular to certain ethnic groups.

The study of cancer risk, cancer risk factors, and intervention efficacy in carriers of disease-susceptibility genes is challenged in several ways. Major challenges arise from the ethical issues in connection with the high cancer risks of study subjects and their offspring, but there are other logistical issues as well. Because the prevalence of carriers of rare mutations is small, large, multisite, expensive studies must be mounted. Cohort studies of carriers and their relatives must be based on random samples of carriers. Retrospective cohort studies of cancer incidence following various interventions in high-risk families must avoid the selection bias of counting those cancers that brought the family to attention. Finally, randomized trials must compare two active treatment arms because allocating an inactive placebo to very high-risk individuals is unethical.

Response Rates and Selection Bias

The problems of incomplete response that plague all epidemiologic studies apply to genetic studies as well, particularly to those attempting to characterize genetic polymorphisms and gene–environment interactions. If response rates differ by ethnicity, comparisons of gene frequency across groups are compromised. Moreover, family-based studies are vulnerable to self-selection by probands with a family history of the disease under study. Potential participants who have reason to believe that they may carry high-risk alleles may refuse to provide DNA because of concerns of employment or medical discrimination for either themselves or their relatives.

Problems Particular to Family Registries

Individuals with cancer sometimes prefer that their relatives not know their disease diagnosis. In addition, genetic testing of multiple family members can raise difficult ethical problems. For example, certain members of multiple-case breast cancer families may prefer not to know their genetic status with respect to BRCA1, but it may be clear from a positive test result in an offspring. More detailed discussion of the issues special to family registries can be found in this monograph in the papers on a breast cancer registry by Hopper et al. *(38)* and on a colorectal cancer registry by Haile et al. *(37)*.

Problems Particular to Certain Ethnic Groups

Members of different ethnic groups have distinct cultural patterns for accepting and handling information on chronic diseases *(39)*. These patterns often restrict the amount of information shared among family members, and they must be accommodated in attempting to gather family histories of disease. A further complication is the likelihood that first-generation immigrants often do not know important details of their relatives' medical

history. In addition, there are difficulties in obtaining interview data and biologic specimens from relatives in other countries. Furthermore, we have found ethnic citizens of the United States of Asian ancestry to be less willing to give blood samples than individuals from other groups.

Ethical, Psychosocial, and Legal Issues

Each of the study designs described above requires informed written consent from participants, but there is uncertainty regarding the proper use of consent in the genetic setting. Two

Table 1. Study designs for identifying and characterizing disease-susceptibility genes

Goal	Design	Strengths	Weaknesses
Identify new disease genes			
	Linkage studies		
	Parametric linkage	Good power when assumptions agree with mechanism generating data	Requires knowledge of mode of inheritance; can be sensitive to violations of assumptions Impaired by presence of phenocopies, etiologic heterogeneity
	Nonparametric linkage	Requires no assumptions about mode of inheritance, marker location	Existing software cannot handle very large families Impaired by presence of phenocopies, etiologic heterogeneity
	Association studies		
	Case-unrelated control	Statistically efficient Straightforward collection of cases and unrelated controls	Confounding by ethnic stratification Less powerful than linkage studies for rare, high-penetrance allelic variants Main effects for gene polymorphism of interest may be misleading in the presence of gene–environment interaction Functional significance of some allelic variants is not known; polymorphisms in noncoding regions may not have functional significance
	Case-related control	Controls for ethnic ancestry	Difficulty recruiting family members for study (incomplete participation) Unavailability of parent controls for older onset diseases In some circumstances, is less efficient than case–control studies for common low-penetrance polymorphisms
Characterize known disease genes			
Estimate allele frequency	Cross-sectional studies	Easy to conduct	Genetic variants associated with poor survival underrepresented in cross-sectional sample
Estimating penetrance	Case-unrelated control	Good power for examining common, low-penetrance polymorphisms	Penetrance may be difficult to assess because mild phenotypes may be undetected in control subjects Incomplete penetrance can affect measures of familial aggregation
	Family cohort	Effective design for assessing penetrance of rare, high-penetrance mutations	Requires knowledge of mode of inheritance Can be biased because of selection factors for probands
Evaluating environmental modifiers of penetrance	Case-unrelated control	Efficiency is good for examining common polymorphisms Acknowledges that not everyone with genotype will manifest phenotype	Confounding by ethnic stratification Lacks power for detecting gene–environment interactions when either polymorphism or environmental factor is low in frequency
	Case-related control	Controls for ethnic ancestry	Family controls overmatched to cases with respect to environmental exposures Lacks power for detecting gene–environment interactions when either polymorphism or environmental factor is low in frequency Can lack power, depending on sample size
Evaluate preventive strategies in genetically susceptible	Randomized trial	Usually free of selection, confouding biases Most valid study design for assessing effect of intervention	Feasibility and power depend on number of carriers, size of risk among genetically susceptible individuals Usually unethical to include a no-intervention control group
	Cohort study	Retrospective cohort studies can be economical	Patients at high risk may selectively seek aggressive interventions Referral centers that see large numbers of high-risk patients may offer only one intervention Power may be inadequate when number of gene carriers is small

issues have been the subject of considerable recent debate. First, must donors give additional written informed consent for the use of archived blood or tissue specimens collected in studies whose purposes differ from those currently planned? The National Bioethics Advisory Commission, a group of lawyers, ethicists, and medical professionals, has recently issued a draft report arguing for tighter controls on stored materials to protect the donors' privacy (posted on the Internet at ⟨www.bioethics.gov⟩). Specifically, new informed consent may be necessary whenever the investigators have not indicated previously to participants the possibility that samples may be used for purposes other than the original study objectives *(40)*. Moreover, donors at this point should be given a chance to ask that their specimens not be used in this or any future research project. If a study involving *de novo* acquisition of specimens might pose more than a minimal risk of harm to the donors, Institutional Review Boards must ensure that the risks are clearly described to donors before consent is sought. In addition, all potential donors should be informed that their biologic materials will be stored for years and that their specimens may subsequently be reanalyzed for currently undetermined research purposes *(41)*.

Second, there has been uncertainty concerning the responsibility of the investigator to notify study participants if genetic testing indicates the presence of a genetic abnormality that confers increased cancer risk. The current view on this issue is that, in obtaining consent for participation in new research, the investigator should indicate exactly what information participants will receive and when this information will be available, i.e., the investigator need not be obligated to reveal genetic findings to participants, but his or her plans to withhold such information must be delineated clearly at the outset *(42,43)*. According to the National Advisory Council for Human Genome Research *(44)*, the decision whether or not to notify participants should depend on the accuracy with which the data and the test predict risk; the efficacy of existing prevention measures; the availability of nondirective education and counseling for family members; and the likelihood of genetic discrimination with respect to health insurance, life insurance, and employment opportunities.

Conclusions

Table 1 summarizes the various designs available to the genetic epidemiologist for gene identification and characterization as well as some of the strengths and weaknesses of these designs. As evident from the table, each design has its drawbacks, and no one design is clearly preferable for achieving a given goal. The classic case–control study based on unrelated individuals with and without the disease of interest tends to be more efficient than family-based, case–control designs that inherently match case patients and control subjects on ethnic ancestry and may also match them on environmental factors. The classic case–control studies also tend to be simpler to conduct and analyze. There is a need for more research on the magnitude of the potential problem of confounding by ethnic admixture that compromises inferences from these studies. This problem can be alleviated somewhat by focusing on polymorphisms with functional significance in the etiology of the disease, combined with statistical methods that adjust for ethnic stratification in the analysis phase of the study.

The workshop included lively discussion of the desirability and feasibility of designing population-based studies to meet all or most of the goals involved in identifying and characterizing genes. In this monograph, Zhao et al. *(45)* present an integrated approach to gene identification and characterization involving ascertainment of families via individuals with cancer (probands) identified in population-based cancer registries. This approach is illustrated here by the population-based family registries for breast cancer *(38)* and colorectal cancer *(37,46)*. When this ascertainment scheme is used, one must decide at the onset on priorities for the extensive task of recruiting extended families, to verify reported disease history, to gather personal data, and to obtain biologic specimens.

Despite the ethical and strategic difficulties in conducting them, adequately controlled observational and randomized studies provide the best mechanisms for progress in evaluating the genotypes predisposing to increased cancer risk, the prevalences of these genotypes, and the lifetime risks of site-specific cancers borne by carriers of these genotypes. It is important that research protocols actively recruit ethnic minorities and, when appropriate, members of both sexes to estimate possible gene–ethnicity and gene–sex interactions.

References

(1) Hartmann LC, Schaid DJ, Woods JE, Crotty TP, Myers JL, Arnold PG, et al. Efficacy of bilateral prophylactic mastectomy in women with a family history of breast cancer. N Engl J Med 1999;340:77–84.

(2) Schaid DJ, Buetow K, Weeks DE, Wijsman E, Guo SW, Ott J, et al. Discovery of cancer susceptiblity genes: study designs, analytic approaches, and trends in technology. Monogr Natl Cancer Inst 1999;26: 1–16.

(3) Collins FS, Guyer MS, Chakravarti A. Variations on a theme: catologing human DNA sequence variation. Science 1997;278:1580–1.

(4) Breslow NE, Day NE. Statistical methods in cancer research. Vol 1. The analysis of case–control studies. Lyon (France): IARC; 1980 IARC Sci Publ No. 32.

(5) Caporaso N, Rothman N, Wacholder S. Case–control studies of common alleles and environmental factors. Monogr Natl Cancer Inst 1999;26:25–30.

(6) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Monogr Natl Cancer Inst 1999;26:31–7.

(7) Schaid DJ. Transmission disequilibrium, family controls and great expectations. Am J Hum Genet 1998;63:935–41.

(8) Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7.

(9) Tu IP, Whittemore AS. Power of association and linkages tests when the disease alleles are unobserved. Am J Hum Genet 1999;64:641–9.

(10) Abel L, Muller-Myhsok B. Maximum likelihood expression for the transmission/disequilibrium test and power considerations [letter]. Science 1997;257:1328–9.

(11) Muller-Myhsok B. Maximum likelihood expression of the transmission/disequilibrium test and power considerations [letter]. Am J Hum Genet 1998;63:664–7.

(12) Le Marchand L. Combined influence of genetic and dietary factors on colorectal cancer incidence in Japanese Americans. Monogr Natl Cancer Inst 1999;26:101–5.

(13) Vionnet N, Stoffel M, Takeda J, Yasuda K, Bell GI, Zouali H, et al. Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. Nature 1992;356:721–2.

(14) Hager J, Hansen L, Vaisse C, Vionnet N, Philippi A, Poller W, et al. A missense mutation in the glucagon receptor gene is associated with non-insulin-dependent diabetes mellitus. Nat Genet 1995;9:299–304.

(15) Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. Am J Hum Genet 1998;62:676–89.

(16) Schubert E, Lee M, Mefford H, Argonza RH, Morrow JE, Hull J, et al. BRCA2 in American families with four or more cases of breast or ovarian cancer: recurrent and novel mutations, variable expression, penetrance and the possibility of families whose cancer is not attributable to BRCA1 or BRCA2. Am J Hum Genet 1997;60:1031–40.

(17) Serova OM, Mazoyer S, Puget N, Dubois V, Tonin P, Shugart YY, et al.

Mutations in BRCA1 and BRCA2 in breast cancer families: are there more breast cancer susceptibility genes? Am J Hum Genet 1997;60:486–95.
(18) Newman B, Mu H, Butler LM, Millikan RC, Moorman PG, King MC. Frequency of breast cancer attributable to BRCA1 in a population-based series of American women. JAMA 1998;279:915–21.
(19) Warner E, Foulkes W, Goodwin P, Meschino W, Blondal J, Paterson C, et al. Prevalence and penetrance of BRCA1 and BRCA2 gene mutations in unselected Ashkenazi Jewish women with breast cancer. J Natl Cancer Inst 1999;91:1241–7.
(20) Price DL, Sisodia SS, Gandy SE. Amyloid beta amyloidosis in Alzheimer's disease. Curr Opin Neurol 1995;8:268–74.
(21) Raunio H, Husgafvel-Pursiainen K, Anttila S, Hietanen E, Hirvonen A, Pelkonen O. Diagnosis of polymorphisms in carcinogen-activating and inactivating enzymes and cancer susceptibility—a review. Gene 1995;159: 113–21.
(22) Caporaso N, DeBaun MR, Rothman N. Lung cancer and CYP2D6 (the debrisoquine polymorphism): sources of heterogeneity in the proposed association. Pharmacogenetics 1995;5 Spec No:S129–34.
(23) Kempkes M, Golka K, Reich S, Reckwitz T, Bolt HM. Glutathione *S*-transferase GSTM1 and GSTT1 null genotypes as potential risk factors for urothelial cancer of the bladder. Arch Toxicol 1996;71:123–6.
(24) Okkels H, Sigsgaard T, Wolf H, Autrup H. Glutathione *S*-transferase mu as a risk factor in bladder tumours. Pharmacogenetics 1996;6:251–6.
(25) Langholz B, Rothman N, Wacholder S, Thomas DC. Cohort studies for characterizing measured genes. Monogr Natl Cancer Inst 1999;26:39–42.
(26) Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams, et al. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 1997;336:1401–8.
(27) Wacholder S, Hartge P, Struewing JP, Pee D, McAdams M, Brody L, et al. The kin-cohort study for estimating penetrance. Am J Epidemiol 1998;148: 623–30.
(28) Gail MH, Pee D, Carroll R. Kin-cohort designs for gene characterization. Monogr Natl Cancer Inst 1999;26:55–60.
(29) Phelan CM, Rebbeck TR, Weber BL, Devilee P, Ruttledge M, Lynch H, et al. Ovarian cancer risk in BRCA1 carriers is modified by the HRAS1 variable number of tandem repeat (VNTR) locus. Nat Genet 1996;12: 309–11.
(30) Taylor JA, Umbach DM, Stephens E, Castranio T, Paulson D, Robertson C, et al. The role of N-acetylation polymorphisms in smoking-associated bladder cancer: evidence of a gene–gene-exposure three-way interaction. Cancer Res 1998;58:3603–10.
(31) Narod SA, Risch H, Moslehi R, Dorum A, Neuhausen S, Olsson H, et al. Oral contraceptives and the risk of hereditary ovarian cancer. N Eng J Med 1998;339:424–8.
(32) Ursin G, Henderson BE, Haile RW, Pike MC, Zhou N, Diep A, et al. Does oral contraceptive use increase the risk of breast cancer in women with BRCA1/BRCA2 mutations more than in other women? Cancer Res 1997; 57:3678–81.
(33) Yang Q, Khoury MJ. Evolving methods in genetic epidemiology. III. Gene–environment interaction. Epidemiol Rev 1997;19:33–43.
(34) Andrieu N, Goldstein AM. Epidemiologic and genetic approaches in the study of gene–environment interaction: an overview of available methods. Epidemiol Rev 1998;20:137–47.
(35) Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case–control studies of candidate genes and gene–environment interactions: basic family designs. Am J Epidemiol 1999;149:693–705.
(36) Goldstein AM, Andrieu N. Detection of interaction involving identified genes: available study designs. Monogr Natl Cancer Inst 1999;26:49–54.
(37) Haile RW, Siegmund KD, Gauderman WJ, Thomas DC. Study-design issues in the development of the University of Southern California Consortium's Colorectal Cancer Family Registry. Monogr Natl Cancer Inst 1999;26:89–93.
(38) Hopper JL, Chenevix-Trench G, Jolley DJ, Venter DJ, Dite GS, Jenkins MA, et al. Design and analysis issues in a population-based, case–control–family study of the genetic epidemiology of breast cancer and the Co-operative Family Registry for Breast Cancer Studies (CFRBCS). Monogr Natl Cancer Inst 1999:26:95–100.
(39) Becker G, Beyene Y, Newsom EM, Rodgers DV. Knowledge and care of chronic illness in three ethnic minority groups. Family Med 1998;30: 173–8.
(40) Bobrow M, Harper P, Harris J, Evans G, Hunt A. Seminar on ethical issues arising from molecular studies in human genetic disease: held under the auspices of the UK Cancer Family Study Group in Manchester-21st May, 1992. Panel discussion. Dis Markers 1992;10:211–28.
(41) Hanning VL, Clayton EW, Edwards KM. Whose DNA is it anyway? Relationships between families and researchers. Am J Med Genet 1993;47: 257–60.
(42) Annas GJ. Privacy rules for DNA databanks. Protecting coded "future diaries." JAMA 1993;270:2346–50.
(43) Andrews LB, Fullarton JE, Holtzman NA, Motolsky AG, eds. Executive summary: assessing genetic risks. Implications for health and social policy. Washington (DC): Natl Acad Press; 1994.
(44) National Advisory Council for Human Genome Research: statement on use of DNA testing for presymptomatic identification of cancer risk. JAMA 1994;271:85.
(45) Zhao LP, Aragaki C, Hsu L, Potter J, Elston R, Malone KE, et al. Integrated designs for gene discovery and characterization. Monogr Natl Cancer Inst 1999;26:71–80.
(46) Siegmund KD, Whittemore AS, Thomas DC. Multistage sampling for disease family registries. Monogr Natl Cancer Inst 1999;26:43–8.

Notes

Supported by Public Health Services grants 5R35CA47448 to Dr. Whittemore (National Cancer Institute), 3RO1NS31964 to Dr. Nelson (National Institute of Neurological Disorders and Stroke), and 5RO1ES08150 to Dr. Nelson (National Institutes of Environmental Health Sciences), National Institutes of Health, Department of Health and Human Services.

We are grateful to Dr. Beth Newman for discussions that helped in the preparation of this paper.

Integrated Designs for Gene Discovery and Characterization

Lue Ping Zhao, Corinne Aragaki, Li Hsu, John Potter, Robert Elston, Kathleen E. Malone, Janet R. Daling, Ross Prentice

Recent advances, including near completion of the human genome map, ever improving high-throughput technologies, and successes in discovering chronic disease-related genes, have stimulated the further development of genetic epidemiology. The primary mission of genetic epidemiology is to discover and characterize genes, whether independent of or interactive with environmental factors, that cause human diseases. To accomplish such a mission, genetic epidemiology needs to integrate both genetic and epidemiologic approaches. One of the challenges facing such an integrated approach is the identification of study designs that are efficient for both gene discovery and characterization. Because designs for gene discovery alone and designs for gene characterization alone have been elaborated in the other two panels, the focus of this paper is to describe those designs that may be useful for discovery and characterization jointly, including case-family and case–control-family designs. Examples of integrated designs are described, and studies of breast cancer conducted at the Fred Hutchinson Cancer Research Center are used for illustration. Finally, related analytic issues are also discussed. [Monogr Natl Cancer Inst 1999;26:71–80]

The ultimate goals of genetic epidemiology include not only the discovery of novel functional genes and investigation of their functional properties—the primary objective of gene discovery—but also the establishment of population allele frequencies and genotype penetrances in relationship to specific human diseases—the primary objective of gene characterization. Complexities of common human diseases, from a genetic perspective, have been described by Schaid in the Gene Discovery Panel *(1)*, earlier by Lander and Schork *(2)*, and by others. These complexities make research more difficult not only in gene discovery but also in gene characterization. To succeed in dissecting complex diseases, one needs to consider a comprehensive approach, based on the analysis of systematically collected data with the use of multiple analytic tools. Earlier in 1997, we described a design framework *(3)* to integrate population-based and family-based designs using a multistage approach *(4–8)*. As a framework, it is inclusive of most study designs used to discover genes, such as affected sib-pairs and highly selected families, as well as of designs to characterize candidate genes, such as population-based case–control and case-relative–control study designs. Because of its inclusive nature, one utility of this framework is to serve as a paradigm for categorizing different study designs.

Because design topics on gene discovery and on gene characterization have been explored fully by investigators in the Gene Discovery Panel and Gene Characterization Panel, respectively, this paper will focus on integrated designs for the purpose of both gene discovery and gene characterization. After describing the need for and challenges to integration, this paper briefly introduces the population-based family-study design framework, including a description of several new results beyond those discussed in our earlier paper *(3)*. We list examples of designs for both gene discovery and gene characterization and then attempt to evaluate critically this framework from various perspectives. For illustration, we describe studies of breast cancer conducted at the Fred Hutchinson Cancer Research Center. Finally, we outline a likelihood framework that may be useful for developing statistical methods, taking advantage of this design framework, for calculating sample sizes and study power, for comparing methods, and for developing robust and efficient methods.

Rationale

The majority of common diseases, especially cancer and coronary heart disease, appears to have a complex etiology, including incomplete penetrance, phenocopies, genetic heterogeneity, gene–gene interactions (epistatic), and gene–environment interactions. Dissecting complex traits calls for an interdisciplinary approach. Recognizing this emerging need, researchers with diverse scientific backgrounds have begun interdisciplinary communication and, in some cases, have collaboratively proposed interdisciplinary studies.

In such research, one of the first challenges is how to design such a study that is efficient and appropriate from both genetic and epidemiologic perspectives. From a genetic perspective, studies should focus on informative samples, improving the chance of finding disease genes. Such considerations often lead geneticists to select families with "unusual" family histories, since such families tend to have an enhanced probability of carrying disease genes. From the perspective of epidemiologists, population-based sampling is key to reliable detection of significant etiologic factors in the study population. Furthermore, under population-based sampling, research results can be generalized to the target population. Motivated by such considerations, epidemiologists tend to design studies that systematically identify case patients and control subjects in population-based, case–control studies. Cohort studies are also motivated by the same considerations but, to keep the discussion focused, will not be elaborated here. Unfortunately, these genetic and epidemiologic

Affiliations of authors: L. P. Zhao, C. Aragaki, L. Hsu, J. Potter, K. E. Malone, J. R. Daling, R. Prentice, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; R. Elston, Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH.

Correspondence to: Lue Ping Zhao, Ph.D., Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N, MW-806, Seattle, WA 98109.

See "Notes" following "References."

© Oxford University Press

perspectives appear to be at extreme ends of the research spectrum, and their integration has thus been a challenge to the development of interdisciplinary work, motivating the development of the population-based family study design framework to be introduced below.

Bearing in mind the genetic and epidemiologic perspectives, one cannot assess a particular study-design choice without appropriate consideration of the corresponding statistical analyses. Table 1 lists some typical statistical analyses performed in epidemiologic and genetic studies. Association analysis, primarily used in epidemiology, assesses association of the disease phenotype with candidate genes as well as with environmental factors. To establish initial evidence to motivate a genetic endeavor, one may conduct an aggregation analysis to examine 1) presence of familial aggregation of disease and, if established, 2) the patterns of familial aggregation as well as 3) evidence in support of any genetic hypothesis. To confirm a genetic hypothesis, one can perform a segregation analysis, providing a tentative estimate for the penetrance and allele frequency of a postulated disease-related gene. Success in these initial analyses leads to a mapping study, via linkage analysis and linkage-disequilibrium analysis. Whereas the goal of linkage and linkage-disequilibrium analysis is used to discover disease loci by studying inherited mutations, loss-of-heterozygosity analysis using tumor tissues can be used to study somatic changes in solid tumors, i.e., to search for genes that may be responsible for initiation, promotion, and progression of solid tumors. Many of these analyses may be integrated into a combined analysis to gain efficiency, as described below.

An immediate benefit of having an integrated design is that the resultant study not only facilitates multiple analyses but may also lead to a gain in statistical efficiency. For example, when linkage and linkage-disequilibrium approaches are combined, the resulting analysis may allow one to detect the position of putative disease genes with improved efficiency. One may also combine linkage-disequilibrium with association analysis so that one can adjust for known etiologic factors and can search for putative factors that independently or interactively contribute to disease. Other benefits of having integrated designs from a statistical perspective are enumerated in the "Methodologic Considerations" section.

Brief Introduction to Population-Based Family Study Designs

We have previously described a general framework for designing population-based family studies *(3)*. Although the framework is general (*see* "Examples of Integrated Design" section), a comprehensive design would have three sampling stages: association and aggregation (A), segregation (S), and linkage (L), as illustrated in Fig. 1. At stage A, a study may adopt, e.g., a case–control design, sampling a group of case patients within a specific time period from a well-defined population, e.g., a population covered by a cancer registry, or the enrollees of a health maintenance organization, and also sampling a group of unaffected individuals from the appropriate comparison population. From each case patient or control subject, researchers gather questionnaire-based information and biologic samples systematically. The primary objective at stage A is assessment of the associations of candidate genes and environmental factors with a disease phenotype, with the secondary objective of assessing familial aggregation of the disease phenotype with the use of family history data. At stage S, the study ascertains prespecified relatives of case patients and control subjects, gathers questionnaire-based information from all participating relatives, and collects biologic samples from a subset of participating relatives. The primary objectives at stage S include (*a*) assessing residual familial aggregation after adjusting for candidate genes and environmental factors as covariates and assessing whether the aggregation pattern supports a genetic hypothesis, then (*b*) describing the penetrance of the putative gene and corresponding allele frequency via segregation analysis, and, finally, (*c*) quantifying the residual familial aggregation after adjusting for the putative gene and those covariates. At stage L, the researcher selects and determines markers in families that appear to have unusual disease patterns, since such families and pertinent family members are likely to carry abnormal alleles. In addition, stage L may include relatives who have not participated in stage S. To take advantage of high-risk family registries, stage L can even include new families that are from the same general population. The primary objective of stage L is to loealize genes, which, independently of or interactively with candidate genes or environmental factors, cause the disease phenotype.

This framework *(3)* has many features noted earlier, some of which are briefly described here. As a framework, it encompasses many designs that are suitable for either discovering genes or for characterizing candidate genes. Combining these two types of designs leads to the development of "hybrid designs" that can be used for both gene discovery and gene characterization. This melding of designs is made possible by multistage sampling. Although each stage has its own primary objectives, data collected from all stages can be analyzed jointly, and the findings are then interpretable in the context of the underlying general population.

While recognizing desirable features of this framework, one

Table 1. Typical statistical analyses performed in genetic and epidemiologic studies

Type of analysis	Symbol	Purpose
Association analysis	A	Testing a direct association of observable covariates with the phenotype
Aggregation analysis	Agg	Establishing familial aggregation of the phenotype
Segregation analysis	S	Evaluating possible penetrance and allele frequency for a putative disease gene via a random effect model
Linkage analysis	L	Testing the presence of cosegregation of markers with putative disease genes
Linkage-disequilibrium analysis	LD	Testing allelic association between markers and putative disease genes
Loss-of-heterozygosity analysis	LOH	Examining genomic instability in abnormal tissues
Combined analysis		Combining two or more of the above analyses into one for possible gain in efficiency, compared with separate analysis

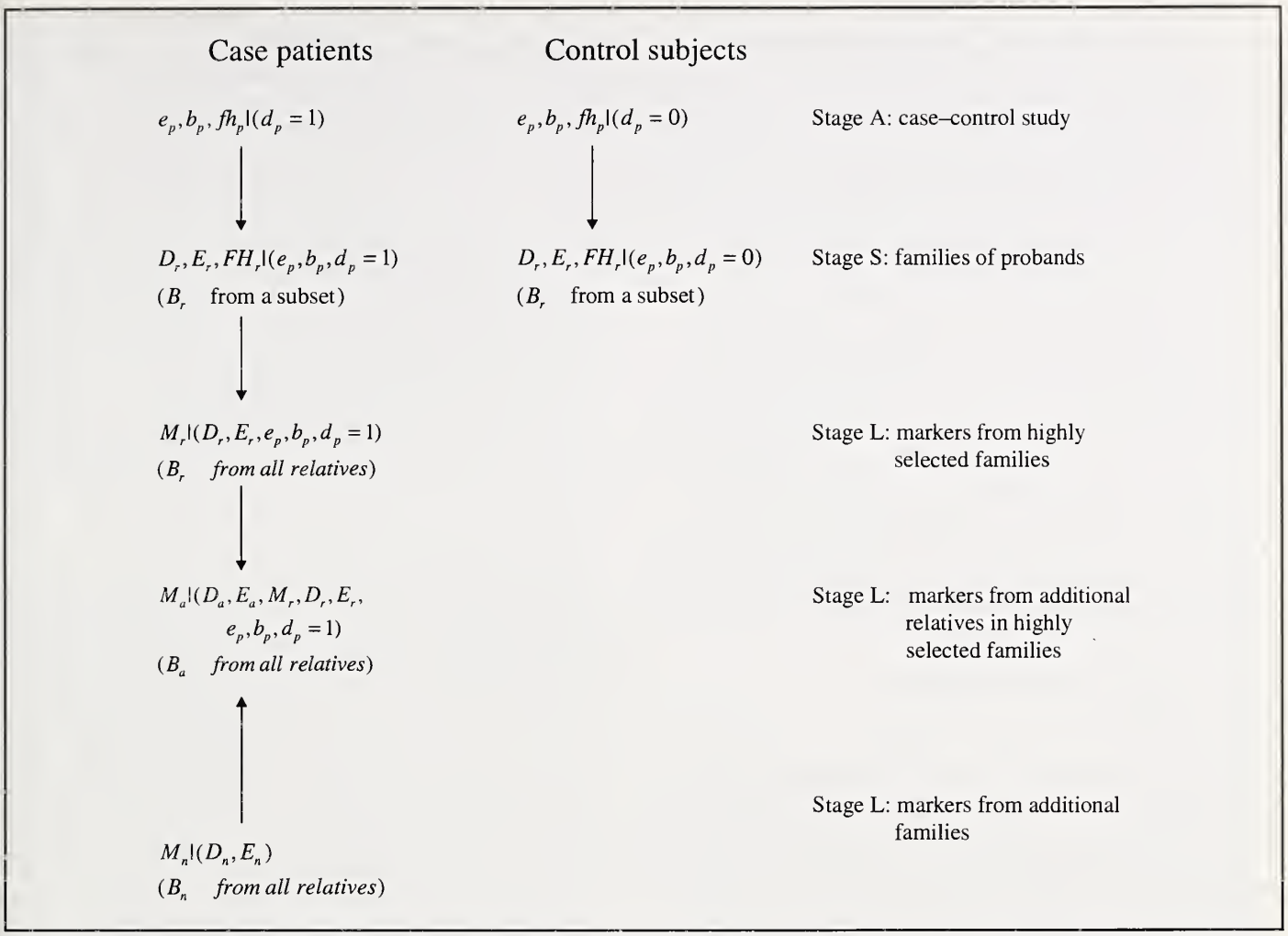


Fig. 1. Flowchart for a three-stage population-based family study design. Subscript p = proband; subscript r = relatives; subscript a = additional relatives; subscript n = new families, lower case letters = a scalar variable; upper case letters = a vector; D or d = disease phenotypes; E or e = environmental factors; B or b = biologic samples; M or m = markers; and FH or fh = self-reported family history.

needs to also bear in mind its limitations. Probably the most significant limitation is that following such a design framework through multiple stages may complicate study conduct and may even encumber studies of gene discovery. To avoid delays in gene discovery, it is necessary to design a protocol that facilitates the quick identification of high-risk families without invalidating the overall study design.

Since our publication on this design framework, we have found some additional related features of interest. First, this framework may be considered as an extension of the epidemiologic design paradigm by integrating family studies into population-based studies (Aragaki, personal communication). This recognition helps summarize the framework in a succinct manner, and it may help epidemiologists appreciate the complexity of conducting population-based family studies. Second, a three-stage study may allow one to discover new genes, to characterize genes, and again to discover additional genes after adjustment for discovered genes, forming a cycle of gene discovery and characterization (*see* Fig. 2). Theoretically, such a cycle allows researchers to uncover more than one gene at a time and poten-

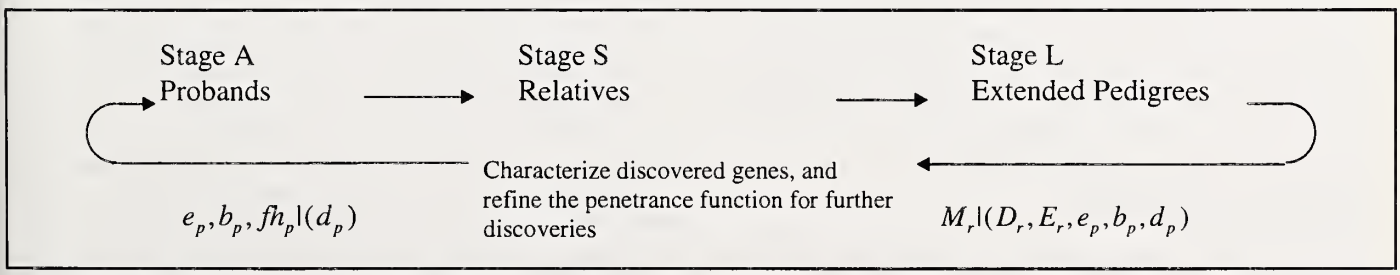


Fig. 2. Schematic flowchart for a three-stage population-based family study. Lower case letters for individual data, upper case letters for all data in families; subscript p = proband; subscript r = relative; d = the disease state; e = environmental covariates; b = biologic samples; fh = self-reported family history; and m = genetic markers.

tially to discover many genetic factors in a single study. Although conceptually appealing, this approach will likely experience limited power at some point of the cycling process, when the majority of the familial aggregation has been explained or when residual familial aggregation is caused by poorly measured environmental factors. Third, we have taken a more detailed look at linkage-disequilibrium analysis and its potential for mapping complex traits. Our experience to date, based on simulated genome scan data (Aragaki, personal communication) *(9)*, supports the idea that the combined analysis of both linkage and disequilibrium signals should be a powerful tool for fine-scale mapping. Unlike linkage analysis, linkage-disequilibrium analysis is preferably carried out on individuals or families that are systematically ascertained from a well-defined population to minimize potential biases as a result of ascertainment or population admixture. In this sense, design considerations for gene discovery and for gene characterization are somewhat convergent. Fourth, the development of a high-density genome map also facilitates a genome scan for loss of heterozygosity. To have an accurate view of the genome-wide pattern of loss of heterozygosity, one generally requires systematic collection of normal and abnormal tissues at various disease stages, calling for special design consideration within this framework.

Examples of Integrated Designs

Case-Only Study

The case-only study design involves sampling population-based cases (or systematically ascertained cases from a well-defined cohort) and may be considered as a study with a single stage A. From each case patient, a tissue specimen is taken, and its genomic DNA is extracted and genotyped for an array of genetic markers. Under typical population-genetic assumptions for the study population, genetic markers in the population may be thought of as independently distributed and thus follow Hardy–Weinberg equilibrium. However, in the presence of linkage-disequilibrium at certain marker loci, the genetic markers from patients would deviate from Hardy–Weinberg equilibrium proportions. Scanning through the human genome, one can test for association between marker alleles and putative disease-related alleles *(10)*.

When solid-tumor tissue samples are collected from patients, both the abnormal part of the tissue and the adjacent normal tissue may be extracted and genotyped for the same genetic markers. A direct comparison of genetic markers between normal and abnormal tissues would reveal the presence or absence of loss of heterozygosity at those loci. The pattern of loss of heterozygosity is indicative of genomic instability that occurred during disease initiation, promotion, and progression. Combining linkage-disequilibrium and loss of heterozygosity, one may be able to detect those genetic alterations that are important as both germline and somatic mutations. Although this design is useful for the stated objectives, it has some limitations in the assessment of gene–environmental interactions (to be discussed in "Discussion and Summary" section).

Case-Relative–Control Study

The case-relative–control study involves sampling population-based case patients and their unaffected relatives as matched control subjects, which again may be thought of as a stage A study. Although biologic samples collected from cases can be used for linkage-disequilibrium and loss-of-heterozygosity analyses as described above, the primary reason for including relative controls is to facilitate an analysis of the association between genetic markers and disease phenotype. Improving on the usual population-based, case–control design, the case-relative–control study is better able to yield unbiased estimates of odds ratios for the association of candidate genes with phenotype because it potentially adjusts for the confounding effects as a result of population admixture or genetic heterogeneity because of different founders *(11,12)* and for shared environmental factors. While appreciating the usefulness of the linkage-disequilibrium analysis, we need to realize that this design, and those described later (two-stage case-family–control and two-stage case–control family studies), are not efficient for linkage analysis. Such a design is justified primarily on the ground of cost efficiency and feasibility for mapping complex traits using association, linkage-disequilibrium, and loss-of-heterozygosity analyses.

Two-Stage Case-Family Study

The case-family study involves sampling population-based case patients at stage A and ascertaining family members (prespecified by degree of kinship), regardless of their disease status at stage S. Because of the potentially large number of relatives in such studies, collection of biologic samples often has to be restricted to case patients and, possibly, affected relatives and selected unaffected relatives. For example, a case-family study of a common disease may sample incident case patients from a well-defined population at stage A and then ascertain all their first-degree relatives, including parents, siblings, children, and spouse(s), at stage S. Even for "common" diseases, the majority of case-relatives are unaffected at the time of the study. Hence, one or more unaffected relatives could be used as control subjects, and the corresponding analyses for a case-family–control study may be performed. If all of the relatives are included, however, one may perform two additional analyses: aggregation analysis and segregation analysis. Through aggregation analysis, one can estimate the magnitude of the familial aggregation of phenotypes and can establish patterns of familial resemblance. Furthermore, a combined association and aggregation analysis, after adjusting for known covariates (candidate genes or environmental factors), may identify residual familial aggregation and may indicate whether or not it supports a specific genetic hypothesis. Once a genetic hypothesis is formulated, one can use the combined association and segregation analysis to estimate the penetrance parameters of potential putative gene(s) and the corresponding allele frequencies. Furthermore, a combined association, segregation, and aggregation analysis may be useful in quantifying residual familial aggregation after adjusting for the putative gene(s), in addition to covariates. An absence of residual familial aggregation would suggest limited additional genetic contribution to the complex disease. Although such a case-family dataset is useful for all of these purposes, our experience to date indicates that a combined association and aggregation analysis, or any other combined analysis that includes those two analyses, may experience nonidentifiability of key parameters in the absence of external information. Furthermore, from an epidemiologic perspective, such a study tends to overmatch on environmental factors (and candidate genes) that are shared within families and, therefore, has reduced power to study environmental factor-related associations. To address this defi-

ciency, one may include population-based control subjects as described in the two-stage case-family control study. From a genetic perspective, however, limited power for linkage and linkage-disequilibrium analysis would be a serious concern. Addressing such a concern motivates the third stage of the data collection to be described below.

Three-Stage Case-Family Study

Following the paradigm outlined above, one may identify those families, satisfying the prespecified sampling criteria, with an "unusual" family history. The study then ascertains all members of those families because those relatives are more likely to carry genetic defects of interest. The purpose at this stage of data collection, known as stage L, is to optimize the efficiency of linkage and linkage-disequilibrium analysis. Also at stage L, one may include additional families that are samples of convenience from, e.g., high-risk family registries, which meet certain sampling criteria and yet have not been part of the first two sampling stages. Although naively pooling such ad hoc samples is inappropriate for association and aggregation analysis, adding such families is valid for linkage analysis, provided that these highly selected families are from the same study population and thus share the same penetrance and allele frequency. This assumption is, unfortunately, not readily verifiable with available data. However, the ultimate proof of a successful linkage analysis is the localization of disease genes.

Two-Stage Case-Family–Control Study

Extending the two-stage case-family study design, the two-stage case-family–control study design collects, in addition to those cases and their prespecified relatives, population-based controls at stage A. The primary reason for including population-based control subjects is to improve the efficiency of assessing either environmental and lifestyle variables or candidate genes in relation to disease phenotype and, more important, to improve the chance of finding genes that may be involved in gene–candidate gene interactions or in gene–environment interactions.

Following the same rationale for extending the two-stage case-family study to the three-stage case-family study, a three-stage case-family–control study adds highly selected families at stage L to optimize the efficiency of performing linkage and linkage-disequilibrium analysis.

Two-Stage Case–Control Family Study

The two-stage case–control family study extends the two-stage case-family–control study design by including prespecified relatives of control subjects at stage S, a design that has also been detailed by Hopper et al. *(13)*. The usually large number of control relatives necessarily limits collection of biologic samples and questionnaire information. In addition, the prevalence of many human diseases may be low among relatives of control subjects, e.g., from 4% to 7% for breast and colorectal cancers. Despite these limitations, control families are valuable for (*a*) assessing familial aggregation of environmental factors or lifestyle variables in the general population, which may be compared with the pattern in families of cases and is informative about the role of potential environmental factors, and (*b*) assessing aggregation and segregation through comparing disease occurrence among relatives of case patients with that among relatives of control subjects, the analytic tools that require fewer assumptions and hence lead to more robust conclusions. In practice, the decision to sample control families has to be made on a case-by-case basis, balancing the information gain against the required data-collection effort and cost.

As described above, one may be interested in adding stage L to the above two-stage case–control family study, resulting in a three-stage case–control family study. In principle, this stage includes high-risk families of case patients, or those of control subjects, or even families from external sources.

Evaluation of Integrated Designs

Integrated designs for gene discovery and for gene characterization represent a shift in the research paradigm. In spite of the motivations listed above, such a shift should be critically evaluated. In this section, our objective is to evaluate integrated designs from genetic, epidemiologic, and analytic perspectives by considering the following questions.

Why would one want to integrate designs for gene discovery and characterization? In contrast to integrated designs, a traditional approach is to design a gene discovery study and, if one or more genes are mapped, to design or use a separate population-based study to characterize them. Naturally, geneticists in the gene discovery study could optimize the study efficiency by using, for example, several large pedigrees or a series of affected relative pairs. Conversely, epidemiologists could follow the lead of mapped genes and could design a separate study to characterize the genes. A successful example is the story of discovering and characterizing the BRCA1 gene. The primary advantages of this traditional approach include 1) the simplicity of designing focused research projects, which ensures the timely discovery of disease genes, and 2) the feasibility of conducting such projects in stages, so that the characterization study is not proposed until it is proven necessary. The disadvantages include 1) the difficulty of assessing the importance of discovered genes in the general population, since assessing penetrance and allele frequency is not feasible with high-risk families; 2) the overly long time lag before mapped genes are characterized, therefore slowing down the translational research to clinical practice or to cancer prevention and control; and 3) inefficiency because genotyped data from highly selected families are not easily used for characterizing genes.

An integrated design may overcome these disadvantages. As noted earlier, most designs for gene discovery and for gene characterization are encompassed by this framework, and the actual choice of a particular design depends on the objectives and the available resources. For example, if discovering genes has a high priority, one could design a three-stage study as follows: At stage A, one would systematically ascertain all cases and administer a simple family-history questionnaire, without collecting either detailed information or biologic samples. On the basis of simplified family-history information, one can ascertain relatives of interest at stage S, and a minimum amount of information is collected. Information on relatives, especially verified disease phenotypes, can now be used for sampling at stage L. Extensive information and biologic samples are gathered from all family members and are used for gene discovery. Once promising genes have been identified, one can now recontact probands and their relatives gathered at stages A and S for characterizing genes.

Why population based? Whereas population-based designs have been extensively used in epidemiology to ensure unbiased

estimation and valid inference, their value is increasingly appreciated in genetic research. For example, if a linkage-disequilibrium analysis is used to map disease genes, it is essential to ensure that the study samples are population based; e.g., haplotypes with the normal and abnormal alleles should be representative of their respective populations. Population-based designs become even more critical to the estimation of allele frequencies in the general population, an important aspect of gene characterization. Furthermore, if there is an interest in gene–environmental interactions, population-based samples become even more important to ensure valid estimation and inference.

Why family based? The essence of gene discovery is to find putative genes that explain the disease aggregation within families; hence, family-based designs are commonly used in genetic research. In contrast, epidemiologic studies generally focus on the direct association of measured genes or environmental factors with disease phenotype and tend to use, instead of family data, independent case patients and control subjects so that no genetic or environmental factors are overmatched between family members. Nevertheless, these contradictory design considerations have been converging in recent years, as family-based, case-relative–control designs have been advocated for gene characterization studies. The primary reason for using a case-relative–control design is to overcome the confounding effects as a result of population admixture, especially for studies in the multiethnic, multiple founder population like the United States.

Is there a single optimal integrated design? An optimal design choice depends on the nature of the disease phenotypes and the underlying disease genes to be discovered. Furthermore, the choice depends on the resources available to investigators. Nevertheless, being guided by the above design framework, investigators could think through issues relating to gene discovery and characterization and could come up with a desirable integrated design for the specific problems to be addressed.

What are the practical utilities of integrated designs? The above design framework may help investigators think through many critical issues for both gene discovery and gene characterization. One important utility is to set research priorities and to allocate research resources. For example, the Cooperative Family Registries of Breast and Colorectal Cancer as well as the Cancer Genetic Network sponsored by the National Cancer Institute represent such long-term resource development activities for cancer genetic research, at least for the next decade. If appropriately designed, such resources could be efficiently used to answer a wide range of research questions. A second utility of this framework is that it facilitates the planning of interdisciplinary research projects, such as program project grants having several components. By using a well-planned research framework, projects within such a program can share all the data being gathered and can perform individual analyses using all available information from all projects to achieve efficiency and population-based interpretation. A third utility is that it helps individual investigators prioritize a long-term research agenda.

Population-Based Family Study of Breast Cancer

In 1993, we (Zhao, Malone, Daling, and Ostrander) embarked on a genetic epidemiology study designed to investigate the independent and interactive contributions of genetic and lifestyle or environmental factors to the familial aggregation of breast cancer in a well-defined population-based series of young case patients with breast cancer and age-matched control subjects. The goals of the study include the conduct of aggregation and segregation analyses that account for exposure to risk factors for breast cancer in both probands and their relatives, the identification and characterization of new candidate genes for breast cancer, the screening of case patients and control subjects for identified susceptibility genes for breast cancer, and the assessment of interaction between environmental and genetic factors. As an example of the proposed design framework described above, this study shows the versatility of population-based studies for accommodating diverse scientific goals.

Original Case–Control Studies

Two population-based case–control studies of breast cancer in young women provided the foundation for this genetic epidemiology study. Both studies, led by Dr. Janet Daling, were designed to evaluate the role of reproductive factors and oral contraceptives in the etiology of breast cancer and used similar methods. For the first study, all incident cases of breast cancer diagnosed from January 1, 1983, through April 30, 1990, among women born after 1944 who were residents of three western Washington counties were ascertained and approached for interview. Case patients were identified through the Cancer Surveillance System, a population-based cancer registry that participates in the Surveillance, Epidemiology, and End Results (SEER) Program[1] of the National Cancer Institute. The Cancer Surveillance System ascertains more than 99% of all incident cancer cases in the 13 western Washington counties. A population-based control group was ascertained with the use of random-digit telephone dialing within each of the three counties from which the cases were drawn. A total of 845 women with breast cancer and 961 control subjects were successfully interviewed (83.2% of eligible case patients and 75.5% of eligible control subjects, respectively). In the second case–control study, all incident cases of breast cancer diagnosed from May 1, 1990, through December 31, 1992, in women under the age of 45 years residing in the three-county area were ascertained and approached for interview. Interviews were completed with 643 women with breast cancer and with 610 control subjects (86.4% and 78.1% of those eligible). Both studies used standardized in-person questionnaires that elicited information on a wide array of risk factors, including reproductive, contraceptive, and menstrual history; lifestyle factors, such as alcohol, smoking, and dietary intake; medical history; family history (enumeration of female relatives, questions on years of birth and death, history of cancer, and age at diagnosis for each relative); and demographics.

Other Ancillary Studies

Case patients and control subjects identified in these two studies have been used for several additional ancillary studies. Currently, we are following all 1288 previously interviewed women who were diagnosed with invasive breast cancer and who agreed to participate in the current study. The case patients (or their proxies) are approached for completion of a questionnaire eliciting data on exposures after diagnosis, treatment history, and recurrences. To date, questionnaires and release forms for tumor access and chart review have been obtained for more than 86% of the case patients. Tumor tissue blocks are obtained by the project pathologist, Dr. Peggy Porter, from cases for

evaluation of potential prognostic factors, including markers of proliferation, tumor suppressers, cell-cycle genes, and genes involved in apoptosis, and medical records are reviewed to evaluate and control for the impact of treatment.

Genetic Epidemiology Study

The previously interviewed case patients and control subjects were recontacted in this study and asked to update and expand previously reported family history information. Female relatives (mothers, sisters, aunts, cousins, grandmothers, or their proxies) of 550 case patients and 550 control subjects (from the second case–control study) were asked to provide data on their own risk factor histories through a telephone interview (5500 interviews completed thus far), and relatives from selected families, such as those with four or more affected members, were asked to provide blood samples.

BRCA1 and BRCA2 mutation analyses are under way, led by Drs. Malone and Ostrander, for defined subsets of the case population, such as women diagnosed before the age 35 years and women with a first-degree family history of breast cancer *(14,15)*. Because these analyses rely on a population-based ascertainment scheme, they avoid the somewhat limited generalizability associated with studies of high-risk women ascertained through genetic clinics or through families accumulated by referral because of their unusual and extreme profiles. In addition, the families of women with a mutation are being characterized to evaluate founder effects within these families. We have also been well positioned to investigate the contributions of a number of other candidate genes, including ATM and the ER gene.

Prior analyses of family history data from our first case–control study revealed that women with a first-degree family history of breast cancer had a 50% lower risk of dying of the disease than did women with no family history of breast cancer or women with a second-degree-only family history of breast cancer *(16)*. As the number of breast cancer cases tested for BRCA1/BRCA2 increases, we have also begun to investigate the relationship of these factors to tumor characteristics and the risk of dying. This unique infrastructure built on a well-defined population-based series of cases and controls offers rich scientific benefits in a number of different directions.

Methodologic Considerations

Framework for Developing Methodologies

One immediate statistical benefit of having a design framework is that it allows a unified statistical approach to the analysis of complex data collected in multiple stages. The following discussion centers on the derivation of the likelihood functions that are useful for developing statistical methodologies.

Likelihood for Data Collected at Stage A

The data (d_p, x_p), where d_p is an indicator for the proband disease status and x_p represents the corresponding covariates from probands, are generally informative about the associations between covariates and phenotype. Following usual statistical considerations, the likelihood function for retrospective data may be constructed as

$$l(d_p, x_p) = \prod_i f(x_i|d_i), \qquad [1]$$

where $f(x_i|d_i)$ denotes an individual contribution from the ith case or control proband. The above likelihood function may take a different form if stage A is a matched case–control study or a cohort study. The likelihood function for linkage-disequilibrium and loss of heterozygosity among cases can be similarly constructed, treating "x_i" as markers.

An important variation of the above likelihood is an extension that incorporates the family history information reported by probands. This additional information can be used to assess familial aggregation of phenotype, in conjunction with the association analysis. The likelihood function may be modified to

$$l(D_r, d_p, x_p) = l(d_p, x_p)\prod_i f(D_{ri}|x_i, d_i), \qquad [2]$$

where D_r represents a vector of phenotypes reported on relatives and represents a likelihood contribution from the family history as reported by a proband. The covariates of relatives can also be incorporated, but they may be of doubtful reliability when reported only by a proband. Note that the upper case letter is generally used here to denote a vector of variables from a family. Note that while the notation "i" is used to denote the ith family, we also use letters "r" and "p" to denote relative and proband, respectively.

Likelihood for Data Collected at Stage S

Data collected from stage S can be used for either 1) association studies via a case-relative control study design; 2) linkage studies via affected sib-pairs; 3) combined association and aggregation analysis; or 4) combined association, segregation, and aggregation analysis. The likelihood functions to be constructed need to be specific to the objectives of the analyses. The likelihood contribution for the association analysis has been comprehensively dealt with by the Gene Characterization Panel and has been described by Witte et al. *(11)*. For linkage analysis, the likelihood function is similar to those described below. Further, the likelihood function for the combined association and aggregation analysis is similar to the one indicated above after including covariates collected from relatives.

Here we describe a likelihood function for combined association, segregation, and aggregation analysis. Let g denote the putative gene that may contribute to familial aggregation. The objective is to estimate its penetrance function $Pr(d|g,x)$ and its allele frequency. Under the typical assumptions for inherited factors, one can derive a joint distribution of putative genes as well as a joint distribution of phenotypes, covariates, and genotype. Then one may construct the following likelihood function,

$$l(D_r, X_r, d_p, x_p) = \sum_{g_p, G_r} l(D_r, X_r, G_r, d_p, x_p, g_p), \qquad [3]$$

where the summation is over all possible genotypes and the likelihood contribution $l(D_r, X_r, G_r, d_p, x_p, g_p)$ on the right-hand side follows directly from the likelihood [2] by including putative genes, using analogous notation.

Likelihood for Data Collected at Stage L

Data to be collected at stage L are generally used for linkage or linkage-disequilibrium analyses. Let us consider the situation in which marker data are collected from only those relatives

identified at stage S. In this case, the likelihood function for the data takes a very simple form and may be written as

$$l(D_r, X_r, M_r, d_p, x_p, m_p) = \sum_{g_p, G_r} l(D_r, X_r, G_r, d_p, x_p, g_p) f(M_r, m_p | G_r, g_p), \quad [4]$$

where the second distribution on the right-hand side of the equation is specified by the linkage and linkage-disequilibrium processes. Note that markers are assumed to be independent of environmental factors given putative genes, an assumption that could be violated in some cases in which markers are directly or indirectly correlated with certain lifestyles.

There is an important modification to the above likelihood function [4] to accommodate additional relatives ascertained at stage L, who have not been part of stages A or S. Suppose that an ad hoc procedure is used to ascertain these additional relatives, i.e., there is no obvious way to quantify the ascertainment rule. The additional data can be included in the above likelihood function as

$$\sum_{g_p, G_r, G_a} l(D_r, X_r, G_r, d_p, x_p, g_p) f(M_r, m_p, m_a | G_r, g_p, G_a) f(G_a | D_a, X_a, G_r, g_p, D_r, d_p), \quad [5]$$

where the third factor is contributed by the additional relatives. When additional families are contributed from external sources and are assumed to arise from the same conceptual population, their genetic markers, along with disease phenotypes and covariates, can also be combined into the above likelihood function. Let the subscript "*o*" denote other families. The likelihood function may then be constructed as

$$\sum_{g_p, G_r, G_a} l(D_r, X_r, G_r, d_p, x_p, g_p) f(M_r, m_p, m_a | G_r, g_p, G_a) f(G_a | D_a, X_a, G_r, g_p) \sum_{G_o} f(M_o | G_o) f(G_o | D_o, X_o), \quad [6]$$

where the additional summation factor is contributed from those external families. The key assumption made in the above likelihood function is that the estimated penetrance and allele frequency from the multistage study are applicable to those additional families. Under this assumption, additional families may be sampled from high-risk family registries, from a genetic counseling clinic, or from other ad hoc sources.

Sample Size and Power

At the planning stage, evaluating sample size is important to establish the feasibility of a proposed study design. After a design with a feasible sample size is chosen, evaluating the study power, after acknowledging the complexities of the disease of interest, becomes critical. Although analytic power and sample size formula would be desirable, the development of such expressions appears to be challenging because of the multiple study stages and various analyses of interest.

Instead, we propose to use a Monte Carlo method to estimate sample size and to evaluate power for studies with specific null hypotheses and specific alternative hypotheses. After considering local resources and budgetary constraints, one may choose a particular design that consists of design choices at stages A, S, and L. Once a design is finalized, one can simulate data following the proposed design protocol. Specifically, the simulation may include 1) families with desired pedigree structures, 2) one or more environmental factors, 3) one or more genetic factors with known positions along the human genome, 4) genetic markers with known density along the human genome, 5) phenotypes that are determined by genetic and environmental factors, and 6) any other ancillary features of proband identification, relative ascertainment, missing data, and measurement errors. The simulated data are analyzed using the methods alluded to above. Repeating the above procedure for a large number (e.g., 1000) of simulation runs allows one to estimate the power of key test statistics for the proposed design. By profiling over a range of sample size scenarios, one can then pick a sample size so that the study will have the desired power. By following a similar procedure, one can evaluate the power of a study given the study design and sample size.

Framework for Comparing Statistical Methodologies

Since the fundamental paper of Elston and Stewart *(17)*, many methods have been developed in statistical genetics. Most of these methods are applicable to genetic analysis on data collected from studies with integrated designs. Within this general design framework, one can compare the relevant methods with the likelihood methods outlined in the "Framework for Developing Methodologies" section.

It is anticipated that the likelihood methods outlined above may turn out to be equivalent to, or similar to, some established methods. In such cases, the general likelihood methods may be thought of as an extension of established methods, in the sense that existing methods are generalized to studies of other designs. Conversely, established methods can be expected, sometimes, to differ from these likelihood methods. In such cases, it is necessary to compare them in term of bias, coverage probability, and power (or efficiency). A favorable result for the likelihood methods helps their establishment. Unfavorable results are also helpful because they would identify areas for improving the likelihood formulation. Theoretically, for designed studies, appropriately constructed likelihood methods should be able to yield estimates of maximal efficiency.

Practicalities of the Likelihood Approach

As a general theoretical paradigm, likelihood has been a foundation for almost all statistical methodologies *(18)*. Without exception, likelihood is the foundation underlying most of statistical genetics *(17)*. Following the same principle, all of the likelihood functions described above serve as an efficient tool for synthesizing the statistical information contained in various types of studies. In applications of the above likelihood functions to actual studies, the likelihood approaches are expected to face challenges, both statistically and computationally.

Statistical challenges are due primarily to the distributional assumptions required by the likelihood approach. For example, to construct a likelihood function (equation 3) for a segregation analysis, one typically assumes conditional independence, i.e., familial aggregation of phenotypes is entirely accounted for by the assumed putative genes. Consequently, the corresponding likelihood function (equation 3) can be written as

$$\sum_{g_p, G_r} l(D_r, X_r, G_r, d_p, x_p, g_p) = \sum_{g_p, G_r} \prod_{j \in \text{Relatives}} f(d_j | x_j, g_j) f(x_p, g_p | d_p), \quad [7]$$

which has been explored in-depth separately *(19)*. In studies of monogenic diseases, this assumption of conditional independence is justifiable and appropriate. The corresponding likeli-

hood methods have been useful in characterizing putative genes and in discovering those genes in past research. However, in studies of complex traits, this assumption becomes problematic, because there are many factors, other than the assumed putative gene, that may cause familial aggregation. The presence of such factors would violate this conditional independence assumption. Naively using the likelihood under conditional independence may lead to 1) biased estimation of segregation parameters *(20)* and 2) inflated type I error because of incorrect estimation of the standard errors *(21–24)*. Overcoming this statistical challenge has motivated recent development of semiparametric methods *(25–28)*.

Computationally, calculating the likelihood function, especially for large pedigrees and for a large number of marker loci, can be extremely challenging. The primary reason is that it requires enumeration over all possible alleles of putative genes on all family members in pedigrees, the computational burden of which could increase exponentially with the number of founders *(29)*. However, multipoint linkage analysis with the use of the Elston and Stewart algorithm *(17)* needs to enumerate all possible phases at multiple marker loci, the computation burden of which could increase exponentially with the number of marker loci. Much of the past research effort aims at circumventing these computational difficulties. Among a large body of literature in this area, the Elston and Stewart algorithm *(17)* is the foundation for the traditional logarithm of the odds (LOD) score method. Simulation-based methods *(30–33)* have been developed. In the area of developing methods for multipoint linkage analysis in medium-sized pedigrees, Idury and Elston *(34)* have been quite successful. Even more recently, we have been developing a semiparametric method for multipoint linkage analysis *(9,28)*. It seems evident that the further development of flexible, robust, and efficient data analysis tools will be an important element in an overall disease gene discovery and characterization program.

Discussion and Summary

The anticipated completion of the entire Human Genome Map by year 2003 and ever improved high-throughput technologies, such as microarray techniques, will facilitate studies for discovering and characterizing disease genes in the postgenome era. In the beginning of this new era, it is timely to have this workshop on study designs. Indeed, this workshop provides an opportunity for us to present our thoughts on this important issue and has further allowed us to critically evaluate those study designs used in practice as well as those that have just been proposed in the workshop.

In this paper, we provide a rationale for integrated study designs and give examples of potential study designs within the framework of population-based family study designs. Furthermore, these designs are evaluated from genetic and epidemiologic perspectives as well as from a practical viewpoint. Designs introduced in this paper serve best as a stepping-stone for genetic epidemiologists to evaluate critically their study design options and to adopt and expand them as appropriate.

Although the above discussion focuses on genetic factors, the theme of the workshop, we should not overlook the important contribution of environmental factors to human diseases. Studying gene–environment interactions is also an important task for genetic epidemiology. When planning a study focusing on gene–environment interactions, one should consider not only those design issues described above but also epidemiologic issues. For example, environmental factors gathered on family members are likely subject to the influence of difference in ages, differences in birth cohorts, and differences in the periods of diagnoses, and thus some designs such as the case-family study may not be entirely appropriate for assessing gene–environment interactions.

Too often, study designs are advocated or dismissed too quickly. To ensure an accurate assessment of the study design, one needs to use statistical methods that require minimum nuisance assumptions and that are efficient for the intended analysis. Only then, with efficient methods, can one begin to assess the validity and the efficiency of the study design. Often, however, only suboptimal statistical methods are developed and are used to evaluate designs. In such cases, the conclusions regarding certain study designs could be misleading.

It is clear that developing efficient statistical methods is critical and is an integrated part of the development of genetic epidemiology. In the past few years, some effort has been made to develop methods for association analysis with correlated family data, for aggregation and segregation analysis using case–control family data and for linkage analysis incorporating environmental factors. How to combine these analyses, in the hope of gaining efficiency, remains an open question as does the incorporation of time-dependent outcomes and time-dependent exposure variables. As expected, developing these methodologies will be one of many active research areas, along with the general development of genetic epidemiology, in years to come.

References

(1) Schaid DJ, Buetow K, Weeks DE, Wijsman E, Guo SW, Ott J, et al. Discovery of cancer susceptiblity genes: study designs, analytic approaches, and trends in technology. Monogr Natl Cancer Inst 1999;26: 1–16.

(2) Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994; 265:2037–48.

(3) Zhao LP, Hsu L, Davidov O, Potter J, Elston RC, Prentice RL. Population-based family study designs: an interdisciplinary research framework for genetic epidemiology. Genet Epidemiol 1997;14:365–88.

(4) Breslow NE, Cain KC. Logistic regression for two-stage case–control data. Biometrika 1988;75:11–20.

(5) Prentice RL. A case–cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 1986;73:1–11.

(6) Langholz B, Thomas DC. Nested case–control and case–cohort methods of sampling from a cohort: a critical comparison. Am J Epidemiol 1990;131: 169–76.

(7) Whittemore AS, Halpern J. Multiphase sampling designs in genetic epidemiology. Stat Med 1997;16:153–67.

(8) Zhao LP, Lipsitz SR. Designs and analysis of two-stage studies. Stat Med 1992;11:769–82.

(9) Zhao LP, Aragaki C, Hsu L, Quiaoit F. Mapping complex traits with single nucleotide polymorphisms. Am J Hum Genet 1998;63:225–40.

(10) Nielsen DM, Ehm MG, Weir BS. Detecting marker-disease association by testing for Hardy–Weinberg disequilibrium at a marker locus. Am J Hum Genet 1999;63:1531–40.

(11) Witte JS, Gauderman WJ, Elston RC, Thomas DC. Asymptotic bias and efficiency in case–control studies of candidate genes and gene–environmental interactions: basic family designs. Am J Epidemiol 1999;149: 693–705.

(12) Hsu L, Zhao LP, Aragaki C. A note on a conditional likelihood approach for family-based association studies of candidate genes. Hum Hered. In press 1999.

(13) Hopper JL, Giles GG, McCredie MR, Boyle P. Background, rationale and protocol for a case–control-family study of breast cancer. Breast 1994;3: 79–86.

(14) Langston AA, Malone KE, Thompson JD, Daling JR, Ostrander EA. BRCA1 mutations in a population-based sample of young women with breast cancer. N Engl J Med 1996;334:137–42.
(15) Malone KE, Daling JR, Thompson JD, O'Brien CA, Francisco LV, Ostrander EA. BRCA1 mutations and breast cancer in the general population. JAMA 1998;279:922–9.
(16) Malone KE, Daling JD, Weiss NS, McKnight B, White E, Voigt LF. Family history and survival of young women with invasive breast carcinoma. Cancer 1996;78:1417–25.
(17) Elston RC, Stewart J. A general model for the genetic analysis of pedigree data. Hum Hered 1971;21:523–42.
(18) Edwards AW. Likelihood. Expanded ed. Baltimore (MD): The Johns Hopkins University Press; 1992.
(19) Zhao LP, Hsu L, Holte S, Chen Y, Quiaoit F, Prentice RL. Combined association and aggregation analysis of data from case–control family studies. Biometrika 1998;85:299–315.
(20) Gail MH, Pee D, Carroll R. Kin-cohort designs for gene characterization. Monogr Natl Cancer Inst 1999;26:55–60.
(21) Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.
(22) Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30.
(23) Prentice RL. Correlated binary regression with covariates specific to each binary observation. Biometrics 1988;44:1033–48.
(24) Zhao LP, Prentice RL. Correlated binary regression using a quadratic exponential model. Biometrika 1990;77:642–8.
(25) Zhao LP. Segregation analysis of human pedigrees using estimating equations. Biometrika 1994;81:197–209.
(26) Zhao LP, Quiaoit F, Hsu L, Aragaki C. An efficient, robust and unified method for mapping complex traits (I): two-point linkage analysis. Am J Med Genet 1998;77:366–83.
(27) Zhao LP, Quiaoit F, Aragaki C, Hsu L. An efficient, robust and unified method for mapping complex traits (II): multipoint linkage analysis. Am J Med Genet 1998;78:48–61.
(28) Zhao LP, Quiaoit F, Aragaki C, Hsu L. An efficient, robust and unified framework for mapping complex traits (III): linkage/linkage-disequilibrium analysis. Am J Med Genet 1999;84:433–53.
(29) Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 1995;57:439–54.
(30) Guo SW, Thompson EA. A Monte Carlo method for combined segregation and linkage analysis. Am J Hum Genet 1992;51:1111–26.
(31) Thompson EA. Monte Carlo likelihood in genetic mapping. Stat Sci 1994; 9:355–66.
(32) Sobel E, Lange K. Metropolis sampling in pedigree analysis. Stat Methods Med Res 1993;2:263–82.
(33) Thomas DC, Cortessis V. A Gibbs sampling approach to linkage analysis. Hum Hered 1992;42:63–76.
(34) Idury RM, Elston RC. A faster and more general hidden markov model algorithm for multipoint likelihood calculation. Hum Hered 1997;47: 197–202.

Notes

[1]Editor's note: SEER is a set of geographically defined, population-based, central cancer registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis, and the NCI makes the data available to the public for scientific research.

Supported in part by Public Health Service grants CA93020, CA33619, CA53996 (National Cancer Institute); AG14358 (National Institute on Aging); GM28356 (National Institute of General Medical Science); and RR03655 (National Center for Research Resources), National Institutes of Health, Department of Health and Human Services.

We thank the participants in the workshop for their helpful comments and discussions.

IV. APPLICATIONS PANEL

Confirmation of Prostate Cancer Susceptibility Genes Using High-Risk Families

Gail P. Jarvik, Janet L. Stanford, Ellen L. Goode, Richard McIndoe, Suzanne Kolb, Mark Gibbs, Leroy Hood, Elaine A. Ostrander

Data from many types of studies support the hypothesis that strong familial components are involved in the etiology of prostate cancer. One way to access such genes is through the study of families with multiple affected family members and, in particular, families with individuals affected comparatively early in life. Several prostate cancer susceptibility loci have been described to date. Confirmation of the linkage and estimation of the proportion of families who are linked in large independent datasets is essential to understanding the significance of susceptibility genes. We explore the methodology used to perform such studies and the factors that can limit the ability to confirm linkage results. We report specifically the example of the HPC1 gene on 1q24–25. [Monogr Natl Cancer Inst 1999;26:81–7]

Prostate cancer is the most common nonsquamous cell cancer and the second leading cause of cancer deaths among men in the United States. In 1998, approximately 184 500 men in the United States will be clinically diagnosed with prostate cancer and, despite advances in treatment and screening, an estimated 39 000 men will die of the disease *(1)*. Incidence rates increased sharply from 1973 to 1992 (particularly among white men), with the increase being attributed to screening and early detection of tumors through the use of testing for prostate-specific antigen (PSA) screening *(2–4)*.

Little is known about the underlying causes of prostate cancer. A number of environmental risk factors have been considered, including industrial exposures, sexually transmitted infectious agents, endocrine factors, and diet. High dietary fat may be associated with invasive prostate cancer *(5–9)*. Evidence also suggests that vitamins D *(10–12)* and E *(13)* may have protective effects. A cause–effect relationship has yet to be demonstrated in either of these cases, however, and no particular environmental factor has been identified as a major cause of the disease.

One of the strongest risk factors for prostate cancer identified to date is a family history of disease, suggesting that genetic factors play a role in prostate cancer susceptibility, etiology, and progression. We discuss here the key factors that suggest a role for prostate cancer susceptibility genes and the likely features of such genes. We review findings that suggest that there is a prostate cancer susceptibility gene at 1q24–25. Finally, we discuss the results in our dataset and others with regard to observed difficulties in confirming these initial findings of linkage.

Epidemiologic Analyses

A number of epidemiologic studies demonstrate familial clustering of prostate cancer, suggesting the existence of underlying genetic predisposition factors. A subset of those studies is summarized in Table 1 *(14–21)*. In each study, family history was evaluated as a risk factor for prostate cancer. When defining exposure as a positive first-degree family history of prostate cancer (father, brother, or sons), the frequency of exposure is about threefold higher among case subjects than among control subjects, with relative risk (RR) estimates across studies ranging from 2.0 to 6.3. Two studies *(22,23)* have suggested that elevated risk may be associated with affected brothers but not with affected fathers, suggesting a possible X chromosome or shared childhood environment risk. Overall, men with two or more first-degree-affected relatives have a fivefold to 11-fold increase in risk compared with men without a family history of prostate cancer *(16,24,25)*. These studies also provide evidence that the risk of prostate cancer in first-degree relatives of probands is higher among patients with early, compared with late, ages at diagnosis.

A genetic contribution to prostate cancer risk is also supported by twin studies. In one such analysis, Grönberg et al. *(26)* analyzed the incidence of disease among 3840 male twin pairs identified from the Swedish Twin Registry linked to the Swedish Cancer Registry. The study population contained 458 cases of prostate cancer. Results suggest a significant influence of genetics on prostate cancer with proband concordance rates of 0.19 for monozygotic twins and 0.043 for dizygotic twins. Subsequent studies *(27,28)* in these and additional cohorts confirmed these initial observations.

Molecular Analysis

A number of loci on different chromosomes have been implicated in the etiology of prostate cancer. These results are not mutually exclusive and suggest that multiple genes contribute to the progression of prostate tumors and that some of these genes may have redundant functions. Studies on allelic loss in sporadic prostate tumors have been a common approach for identifying regions likely to contain prostate cancer genes [e.g. *(29–31)*].

Affiliations of authors: G. P. Jarvik (Department of Medicine, Division of Medical Genetics), E. L. Goode (Department of Epidemiology, School of Public Health and Community Medicine), R. McIndoe, L. Hood (Department of Molecular Biotechnology), University of Washington Medical Center, Seattle; J. L. Stanford, S. Kolb (Division of Public Health Sciences), M. Gibbs, E. A. Ostrander (Clinical Research Division), Fred Hutchinson Cancer Research Center, Seattle.

Correspondence to: Elaine A. Ostrander, Ph.D., Clinical Research Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., D2-190, Seattle, WA 98109-1024 (e-mail: eostrand@fhcrc.org).

See "Notes" following "References."

© Oxford University Press

Table 1. Epidemiologic studies of prostate cancer in relation to family history of prostate cancer

Study, y (reference No.)	No. of subjects		Positive family history (%)		Estimated relative risk*
	Case subjects	Control subjects	Case subjects	Control subjects	(95% CI or *P*)
Krain, 1974 *(14)*	221	221	6.0	1.0	6.3 (*P*<.01)
Fincham et al., 1990 *(15)*	382	625	15.0	5.0	3.4 (2.2–5.3)
Steinberg et al., 1990 *(16)*	691	640	15.0	8.0	2.0 (1.2–3.3)
Spitz et al., 1991 *(17)*	378	383	13.0	5.7	2.4 (1.3–4.5)
Hayes et al., 1995 *(18)*	972	1294	6.9	2.3	3.2 (2.0–5.0)
Whittemore et al., 1995 *(19)*	1500	1581	13.0	5.0	2.5 (1.9–3.3)
Lesko et al., 1996 *(20)*	563	703	17.2	8.3	2.2 (1.5–3.0)
Ghadirian et al., 1991 *(21)*	640	639	14.7	5.0	3.2 (2.1–5.1)

*CI = confidence interval.

Thus far, loss of heterozygosity (LOH) in prostate tumors has been documented for at least 11 different chromosomal arms. Many of these regions have been identified repeatedly, such as chromosomes 8 and 17 *(32,33)*. Chromosome 17 is of interest because it carries a number of genes postulated to have a role in cancer, including the p53 and BRCA1 genes. Of interest, families that carry BRCA1 mutations have been reported to be at increased risk for prostate cancer (RR 3.33; 95% confidence interval = 1.8–6.2) *(34)*. Analyses of prostate cancer cases drawn from a population-based, case–control study, however, revealed little evidence that BRCA1 is a common prostate cancer susceptibility gene *(35)*. In 49 prostate cancer cases who had either been diagnosed early in life (<53 years), were diagnosed before age 65 years, and had first-degree family history of breast cancer or an early onset family history of prostate cancer, only one germline BRCA1 mutation was observed. There is a likely possibility that this single observation was coincidental, as was the delAG1285 mutation that has been frequently observed in the population. This finding suggests that LOH observed on chromosome 17q in many prostate tumors is likely to be unrelated to the presence of germline mutations in BRCA1.

Other techniques that highlight regions of interest include chromosome transfer studies and comparative genome hybridization that implicate chromosomes 2, 12, 11, and 8 *(36–39)*. The data from chromosome 8 are particularly interesting because LOH studies also implicate chromosome 8. However, exhaustive linkage analysis of chromosome 8 by a variety of methods has yet to reveal a gene important in prostate cancer predisposition *(40–41)*.

The identity, penetrance, and role in cancer susceptibility of each of the above postulated genes remain unknown. Two points are clear, however. First, whereas the location and number of the underlying prostate cancer susceptibility genes remain elusive, collectively these studies argue that tumor suppression is a key function of at least some prostate cancer genes [reviewed in *(42)*]. Second, many of these genes are likely to be more important in prostate cancer progression and metastasis and do not represent early events in the development of prostate cancer. For the mapping of susceptibility genes, discussed below, analyses of families with a high incidence of prostate cancer are needed.

Complex Segregation Analysis

Complex segregation analysis (CSA) is useful for testing the hypothesis that a mendelian locus substantially contributes to the risk of a disease. CSA can give information about the likely mode of inheritance and penetrance of prostate cancer susceptibility genes. It can also suggest subsets of families, or the characteristics of men within families, whose disease is most likely because of mutations in an inherited susceptibility gene. CSA cannot, however, reliably distinguish the number of susceptibility loci nor can it provide information about the location of susceptibility genes. CSA can provide estimates of the overall frequency of mutant alleles with a specific set of characteristics, the penetrance of these alleles, and other parameters of the transmission models that can be used in linkage analyses.

Several major CSAs of prostate cancer have been undertaken. The first, undertaken by Carter et al. *(24)* analyzed data on 691 prostate cancer families, originally ascertained from consecutive probands (mean age, 59.3 years) undergoing radical prostatectomy for primary clinically localized prostate cancer at The Johns Hopkins University Hospital, Baltimore (MD) between 1982 and 1989. These data showed that early age of diagnosis and multiple affected family members were the strongest predictors of risk in relatives. Their analysis also suggested that autosomal dominant inheritance of a rare, highly penetrant allele(s) best accounted for the distribution of prostate cancer in the families studied, with carriers having an 88% cumulative risk of prostate cancer by age 85 years compared with only 5% cumulative risk for noncarriers. These investigators proposed that hereditary prostate cancer gene(s) account for about 43% of prostate cancer with diagnosis before age 55 years and about 9% of total prostate cancer.

Similar results have been obtained by Schaid et al. *(43)* with the use of a family-history survey of cancer conducted for 5486 men, including 4288 probands (aged 38–84 years) who underwent a radical prostatectomy for clinically localized prostate cancer at the Mayo Clinic between 1966 and 1987. This study proposed a population frequency of 0.006 for a susceptibility gene and a risk of 89% by age 85 years for carriers. This study, however, suggested that autosomal dominant genes accounted for a higher proportion of all prostate cancer—approximately 68% before age 60 years.

An additional segregation analysis was performed on a population-based sample of 2857 nuclear families ascertained through an affected father diagnosed with prostate cancer in Sweden between 1959 and 1963 *(44)*. This analysis contrasts with the studies by Carter et al. *(24)* and Schaid et al. *(43)*, who both studied hospital-based series of men whose disease was treated with prostatectomy. Selection of probands by prostatectomy may bias selection toward men with less aggressive dis-

ease, since metastatic disease is unlikely to be treated by prostatectomy. Grönberg et al. *(44)* reported that familial clustering of prostate cancer was best explained by dominant alleles with a population frequency of 0.0167 and a lifetime penetrance of 63%, slightly less than the penetrance proposed by Carter et al. Nevertheless, the overall consistency in segregation analyses derived from very different populations provide strong evidence for highly penetrant autosomal dominant prostate cancer susceptibility genes.

Putative Prostate Cancer Locus at 1q24–25

In 1996, the location of a putative prostate cancer susceptibility locus (termed HPC1) on chromosome 1q24–25 was suggested by linkage analysis of 91 high-risk prostate cancer families from the United States and Sweden *(40)*. A maximum multipoint logarithm of the odds (LOD) score under the assumption of heterogeneity LOD (HLOD) = 5.43 (the proportion of families linked, $\alpha = 0.34$), with the postulated locus being close to marker D1S422. This study concluded that about 34% of inherited prostate cancer could be attributed to an as yet unknown gene at 1q24–25. The above conclusions were strongly supported by a nonparametric analysis using the program GENEHUNTER *(45)* in which highly significant results were obtained for multiple markers in this region.

Although the initial report did not suggest any subgroup of families that were more or less likely to be linked, a strong age of diagnosis effect, with nearly all linked families having an average age of diagnosis of less than 65 years has since been noted. Speeifieally, Grönberg et al. *(46)* found that the 40 North American families with an average age of diagnosis less than 65 years yielded a multipoint LOD score of 3.96, whereas the 39 families with an older age at diagnosis yielded a LOD score of −0.84. Assuming heterogeneity, the proportion of families linked was estimated as 66% for the 14 families with the earliest average ages at diagnosis but decreased to 7% for the families with the oldest ages at diagnosis. The majority of families contributing to the positive LOD score in each case were Caucasian. Grönberg et al. *(46,47)* also reported that high-grade cancers (grade 3) were more common in potentially linked families as was advanced-stage disease. These latter two points have been the topic of some dispute, however, with claims that the comparison of characteristics between putatively linked and unlinked families is faulty because of the use of an incorrect reference population *(48)*.

It should be noted that conclusions regarding clinical characteristics of linked families are virtually impossible to make with any certainty until the HPC1 gene is cloned. The presence of a modestly positive LOD score, or even completely shared haplotypes from the chromosomal region in question among affected individuals, does not guarantee that the family is "linked," since such segregation will occur fortuitously some percentage of the time. Conversely, there will be families with low LOD scores but who have a mutation in the gene in question. The LOD scores may be low beeause of phenocopies, i.e., men who have acquired the same disease for "sporadic" reasons unrelated to germline mutations for the gene in question. In addition, because of variable penetrance of mutations, there will be families in which carriers of the mutation develop the disease at a point later in life then predicted by the statistical models or not at all. Therefore, conclusions regarding clinical characteristics of "linked families" will be best drawn from analyses of families who are found to carry true germline mutations, following characterization of the causative gene.

Shortly after the mapping of HPC1, we described an analysis *(49)* of 49 high-risk prostate cancer families for linkage to 1q24–25 with the use of 10 highly polymorphic markers spanning the proposed locus. These 49 families were representative of our families as a whole, which were collected from the Prostate Cancer Genetic Research Study (PROGRESS), a Seattle-based study initiated in July 1995 to identify familial clusters of prostate cancer and, ultimately, inherited susceptibility genes. A toll-free number (1-800-777-3035) was established and distributed via national advertising, media events, and mailings to support groups and urologists to identify potential participants for their suitability. Informed, written consent was obtained from each study participant. The study and its consent and medical record release forms were approved by the Institutional Review Board of the Fred Hutchinson Cancer Research Center. Families were selected for participation on the basis of the number of first-degree relatives diagnosed with prostate cancer, the age at diagnosis of the affected relatives, and the number of living affected relatives from whom blood samples could be obtained. In our first set of 49 families, we observed that the characteristics of the families analyzed compared closely with those described by Smith et al. *(40)*. These PROGRESS families averaged 4.4 men affected per family (range, 3–9) and had an average age at diagnosis of 65.9 years. Forty-six of the 49 PROGRESS families were Caucasian.

Data were analyzed by use of parametric and nonparametric methods. Two models were used for the parametric LOD score method. The first model ("Hopkins model" or "Model A") was identical to that reported by Smith et al. *(40)*, and the second ("Seattle model" or "Model S1") *(49)* was devised by us and based on the age-dependent penetrance values that resulted from the segregation analysis reported by Carter et al. *(24)*. In addition, to protect against falsely rejecting linkage secondary to model misspecification, we also performed a nonparametric multipoint analysis with the use of the GENEHUNTER program *(45)*. In all cases, we observed no evidence for linkage in our dataset in either two-point analyses (using any of 10 markers in the region) or in the nonparametric analyses. No evidence for linkage was observed in a heterogeneity analysis using HOMOG.

More recently, we have expanded our dataset to include 150 families (Table 2; 144 Caucasian and six non-Caucasian) and undertaken a closer examination at this locus with the use of three linkage models: the Seattle model *(49)*, the model described by Smith et al. *(40)*, and a model used in a recent linkage analysis published by Grönberg et al. *(46)*. These complete results *(50)*, including features of the families analyzed, are similarly negative with no evidence for linkage with a subset of families at 1q24–25. Table 3 presents the results from the 150 multiplex prostate cancer families described in Table 2 for one marker, D1S518, which defines the HPC1 locus. The results are highly negative with all models and all values of θ, as well for all other markers analyzed *(50)*.

One consideration in this analysis is the degree to which the families analyzed mimic those described by Smith et al. *(40)*. The mean age of diagnosis of the genotyped affected men is 65.9 years, and the average number of affected men per family is 4.2 (Table 3). These figures are comparable to that of Smith et al.

Table 2. Characteristics of 150 prostate cancer families from the PROGRESS study*

Characteristic	Total group	Caucasian		Non-Caucasian	
		≤65 y	>65 y	≤65 y	>65 y
No. of affected men	637	275	328	26	8
No. of genotyped affected men	509	219	263	20	7
Average No. of affected men per family (range)	4.2 (3–9)	4.3 (3–9)	4.1 (3–8)	6.5 (4–9)	4.0 (3–5)
Mean age at diagnosis of genotyped affected men per family (range)	65.9 (50.8–78.0)	62.0 (50.8–66.0)†	69.5 (65.0–78.0)	62.2 (57.7–63.7)	69.3 (67.0–71.7)

*PROGRESS = Prostate Cancer Genetic Research Study.
†Men genotyped at 1q24 are not identical to the overall set of affected men samples.

Table 3. LOD* scores for 150 prostate cancer families using three transmission models at 1q for marker D1S518

Model (reference No.)	Recombination fraction							
	0.00	0.02	0.04	0.08	0.10	0.20	0.30	0.40
S1 *(49)*	−30.98	−24.28	−19.41	−12.66	−10.24	−3.25	−0.74	−0.08
A *(40)*	−84.51	−50.94	−37.47	−22.72	−18.04	−5.65	−1.38	−0.17
B *(46)*	−12.46	−10.25	−8.51	−5.92	−4.93	−1.84	−0.54	−0.09

*LOD = logarithm of the odds.

(40) in the description of families collected by investigators at The Johns Hopkins University.

Locus Heterogeneity and Power to Detect Linkage

The inability to robustly replicate the findings of Smith et al. *(40)* suggests that locus heterogeneity issues for HPC1 are still not well understood. There may be diagnostic differences between datasets or it may be that a prostate cancer gene at 1q24–25 accounts for significantly less than 34% of disease in "high-risk" families nationwide. How much less is unclear. In our dataset, we have undertaken an analysis to estimate the power to detect linkage. Conservative computations suggest a high power to detect a hereditary prostate cancer (HPC) locus (or loci) by use of a multistep analysis strategy in this large dataset. Power calculations for two-point (one marker and HPC) LOD method analyses were done with 1000 replications with the use of the computer package SIMLINK *(51,52)*. The simulations computed the probability of detecting certain LOD scores, given the autosomal dominant model of HPC transmission discussed above, the minimum pedigree structures, and the DNA samples available. The computations presented are conservative in several ways. They used only the pedigree structure that was available in the time of initial ascertainment. Much more data have been collected on many families in the intervening months. The power analysis also assumed a polymorphism information content (PIC) of 0.7 for the marker loci. A marker with a PIC of 0.8, as several of the markers in the HPC region have, would confer additional power. Another consideration is that, in the case of HPC1, we are conducting a confirmation study with dense markers and, therefore, expect θ to be small. Table 4 calculates power for $\theta = 0.00$ and 0.05. For the HPC1 region, the power analysis for $\theta = 0.00$ is reasonable.

Finally, at the time power analyses were completed, only 114 families were collected, half of which had a mean age of familial prostate cancer less than 65 years and half of which had a mean age of 65 years or greater. The most recent molecular analyses were done with 150 families, approximately half of whom have a mean age under 65 years, using the age-dependent penetrance model (Seattle model) we have described previously. Despite these caveats, power analysis of 114 families indicates acceptable power to detect linkage, even if a locus accounted for less than 35% of prostate cancer in the dataset. Application of these findings argues that, at least in this dataset, significantly less than 35% of families are linked to 1q24–25.

Table 4. Power analysis of 114 Caucasian families with mean age of diagnosis <65 years*

% linked <65 y	% linked total dataset	θ	Expected LOD	Probability LOD > 1	Probability LOD > 1.5	Probability LOD > 2	Probability LOD > 3
100	50	0	11.448	1.000	1.000	1.000	1.000
		0.05	7.264	0.999	0.996	0.993	0.975
90	45	0	7.264	0.999	0.996	0.993	0.975
		0.05	4.751	0.988	0.956	0.905	0.816
80	40	0	4.581	0.988	0.952	0.890	0.789
		0.05	3.132	0.929	0.832	0.684	0.495
70	35	0	2.741	0.908	0.744	0.590	0.399
		0.05	1.915	0.749	0.542	0.353	0.202

*LOD = logarithm of the odds.

Comparative Analyses

The strong HPC1 linkage observed by Smith et al. *(40)* has proven challenging to replicate in other datasets, with mixed results generated by seemingly similar sets of families. By applying nonparametric analyses with the use of GENEHUNTER to studies of families with significant numbers of affected young men, data from Cooney et al. *(53)* and Hsieh et al. *(54)* weakly confirm the observations of Smith et al., with nonparametric linkage scores of 1.72 and 1.83, respectively, and P values of .045 and .036. The significance of NPL scores with P values in the .03–.05 range for a confirmation result is debatable. In a detailed discussion on this topic, Lander and Kruglyak *(55)* have suggested that a nominal P value of .01 should be required for confirmatory evidence of linkage, since such studies still involve searching over an interval. That P values <.05 were seen in two distinct datasets, however, strengthen the argument for confirmation in these datasets. The percentage of linked families in the overall datasets cannot be estimated from these analyses but will be much less than 34%.

In contrast to the above, Eeles et al. *(56)* and Berthon et al. *(57)* found no evidence for linkage at this locus. Eeles et al. reported an NPL score of 0.72 with a P value of .22 in 35 families with four affected men. In these same families, the HLOD = 0.20, with α = 20%. The HLOD for the larger dataset of 136 families was 0.01, α = 4%, suggesting that only 4% of the families were linked to the HPC1 locus. Virtually identical negative data have been presented by Thibodeau et al. *(58)*.

Effects of Heterogeneity

There is currently linkage evidence for four familial prostate cancer loci *(40,57,59–61)*. Because these loci do not account for more than one half of the families reported, other loci are likely to exist. Locus heterogeneity dramatically influences the power to map. This can be seen by examining the fall in expected LOD score as heterogeneity increases. Table 5 demonstrates the tradeoff between sample size, heterogeneity, and power in the simulation studies for our late-diagnosis families. Total sample sizes of 116 and half that, 58 families, are examined. It can be seen that increasing the proportion of families linked to the locus of interest from 60% to 70% or from 70% to 80% results in a nearly 50% drop in the sample size for required equivalent power. This demonstrates the advantage obtained if homogeneous subsets can be constructed for linkage testing. For prostate cancer, families have been divided by age of diagnosis, ethnicity, and density or severity of disease. Homogeneous subsets may also be improved by stratifying on the LOD score at other putative loci. Alternately, linkage methods in which two or more loci may be jointly considered may be advantageous when locus heterogeneity is suspected within a sample.

The problem of heterogeneity also sheds light on how different samples, ascertained in different ways, may enrich their collection for certain clinical features and, therefore, may have different mixes of susceptibility loci and would result in the differential ability to map each locus between samples. This is the reason that the characteristics of samples must be compared when linkage found in one dataset cannot be confirmed in another dataset.

Table 5. Caucasian families with mean age of diagnosis for genotyped affected men >65 years*

No. of families	Linked, %	θ	Max LOD	Probability LOD > 1	Probability LOD > 2	Probability LOD > 3
116	90	0	9.362	1.000	0.999	0.998
		0.05	6.124	1.000	0.977	0.931
116	80	0	5.915	1.000	0.976	0.923
		0.05	4.030	0.981	0.850	0.699
116	70	0	3.585	0.971	0.762	0.578
		0.05	2.441	0.876	0.515	0.290
116	60	0	2.064	0.796	0.402	0.212
		0.05	1.443	0.643	0.220	0.081
58	100	0	7.071	0.999	0.995	0.972
		0.05	4.608	0.992	0.932	0.770
58	90	0	4.608	0.992	0.932	0.770
		0.05	3.127	0.922	0.761	0.473
58	80	0	3.100	0.912	0.730	0.451
		0.05	2.111	0.754	0.472	0.221
58	70	0	1.833	0.663	0.413	0.161
		0.05	1.294	0.510	0.240	0.077

*LOD = logarithm of the odds.

Confounding Factors

We believe that the two factors that most likely complicate locus heterogeneity issues in the mapping of prostate cancer genes are PSA testing and race. PSA came into common usage as a diagnostic tool in 1988. Because different datasets have collected men spanning different diagnosis years, the proportion of men with less aggressive disease who were diagnosed by PSA and categorized simply as "affected" likely differs (perhaps significantly) between datasets and within individual strata. A weighting factor based on the level of PSA in the linkage analyses may shed light on this issue, as may stratification by tumor stage and grade.

Whether or not race is an issue for HPC1 analysis remains unclear. Only two families in the Seattle dataset were African-American. It is clear that prostate cancer is more frequent among U.S. blacks than whites *(62)* and that it is lowest in Asian men. Although the majority of families contributing to the HPC1 linkage results are Caucasian in all datasets, in the work described by Cooney *(53)*, several African-American families contributed to the observation of linkage, with a maximum NPL Z score at marker D1S158 of 1.39 (P = .0848). There was also an overrepresentation of African-American families contributing to the positive LOD score in the original report of Smith et al. *(40)* as well, generating the hypothesis that the gene at 1q24–25 may be more important for prostate cancer susceptibility in African-American than in Caucasian families. In part, because the results in individual families are not strong enough to conclude linkage of that family to the 1q24–25 region, our understanding of any differential effect of the HPC1 locus awaits larger collections of non-Caucasian families.

Conclusions

The confirmation of the HPC1 locus has not been uneventful. The mixed results found by several groups who have tried to confirm this linkage are likely the result of population differences in the samples and a lower proportion of families being linked than was originally estimated. The mixed results do not necessarily indicate that this linkage is a false positive, rather that the HPC1 gene may be differentially important for families with different epidemiologic or clinical characteristics. It remains for the HPC1 gene to be cloned and sequenced before most families whose disease is positively because of mutations in HPC1 can be unambiguously identified.

The confirmation of this and of future loci will be aided by increasing homogeneity in subgroup analysis, which relies on better characterization of the high-risk families. For these reasons, we plan to continue collection of extensive medical and family history data on all affected and unaffected men in PROGRESS families and increase our ascertainment of non-Caucasian families.

References

(1) Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics, 1998. CA Cancer J Clin 1998:48:6–29.

(2) Merrill RM, Brawley OW. Prostate cancer incidence and mortality rates among white and black men. Epidemiology 1997;8:126–31.

(3) Devesa SS, Blot WJ, Stone BJ, Miller BA, Tarone RE, Fraumeni JF. Recent cancer trends in the United States. J Natl Cancer Inst 1995;87: 175–82.

(4) Parker SL, Tong T, Bolden S. Cancer statistics. Can Cancer J Clin 1997; 47:5–27.

(5) Giovannucci E, Rimm EB, Colditz GA, Stampfer MJ, Ascherio A, Chute CC, et al. A prospective study of dietary fat and risk of prostate cancer. J Natl Cancer Inst 1993;85:1571–9.

(6) West DW, Slaattery ML, Robison LM, French TK, Mahoney AW. Adult dietary intake and prostate cancer risk in Utah: a case–control study with special emphasis on aggressive tumors. Cancer Causes Control 1991;2: 85–94.

(7) Ross RK, Shimiuzu H, Paganini-Hill A, Honda G, Henderson BE. Case–control studies of prostate cancer in blacks and whites in southern California. J Natl Cancer Inst 1987;78:869–74.

(8) Kolonel LN, Yoshizawa CN, Hankin JH. Diet and prostatic cancer: a case–control study in Hawaii. Am J Epidemiol 1988;127:999–1012.

(9) Dijkman GA, Debruyne FM. Epidemiology of prostate cancer. Eur Urol 1996;30:281–95.

(10) Ingles SA, Ross RK, Yu MC, Irvine RA, La Pera G, Haile RW, et al. Association of prostate cancer with genetic polymorphisms in vitamin D receptor and androgen receptor. J Natl Cancer Inst 1997;89:166–70.

(11) Taylor JA, Hirvonen A, Watson M, Pittman G, Mohler JL, Bell DA. Association of prostate cancer with vitamin D receptor gene polymorphism. Cancer Res 1996;56:4108–10.

(12) Morrison NA, Qi JC, Tokita A, Kelly PJ, Crofts L, Nguyen TV, et al. Prediction of bone density from vitamin D receptor alleles. Nature 1994; 367:284–7.

(13) Heinonen OP, Albanes D, Virtamo J, Taylor PR, Huttunen JK, Hartmen AM, et al. Prostate cancer and supplementation with α-tocopherol and β-carotene: incidence and mortality in a controlled trial. J Natl Cancer Inst 1998;90:440–6.

(14) Krain LS. Some epidemiologic variables in prostatic carcinoma in California. Prev Med 1974;3:154–9.

(15) Fincham SM, Hill GB, Hanson J, Wijayasinghe C. Epidemiology of prostatic cancer: a case–control study. Prostate 1990;17:189–206.

(16) Steinberg GD, Carter BS, Beaty TH, Childs, Walsh PC. Family history and the risk of prostate cancer. Prostate 1990;17:337–47.

(17) Spitz MR, Fueger JJ, Babaian RF, Newell GR. Vasectomy and the risk of prostate cancer [letter]. Am J Epidemiol 1991;134:108–9.

(18) Hayes RB, Liff JM, Pottern LM, Greenberg RS, Schoenberg JB, Schwartz AG, et al. Prostate cancer risk in U.S. blacks and whites with a family history of cancer. Int J Cancer 1995;60:361–4.

(19) Whittemore AS, Wu AH, Kolonel LN, John EM, Gallagher RP, Howe GR. Family history and prostate cancer risk in black, white, and Asian men in the United States and Canada. Am J Epidemiol 1995;141:732–40.

(20) Lesko SM, Rosenberg A, Shapiro S. Family history and prostate cancer risk. Am J Epidemiol 1996;144:1041–7.

(21) Ghadirian P, Cadotte M, Lacroix A, Perret C. Family aggregation of cancer of the prostate in Quebec: the tip of the iceberg. Prostate 1991;19: 43–52.

(22) Narod SA, Dupont A, Cusan L, Diamond P, Gomez JL, Suburu R, et al. The impact of family history on early detection of prostate cancer. Nature Med 1995;1:99–101.

(23) Monroe KR, Yu C, Kolonel LN, Coetzee GA, Wilkens LR, Ross RK, et al. Evidence of an X-linked or recessive genetic component to prostate cancer risk. Nat Med 1995;1:827–9.

(24) Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC. Mendelian inheritance of familial prostate cancer. Proc Natl Acad Sci U S A 1992;89: 3367–71.

(25) Carter BS, Bova GS, Beaty TH, Steinberg GD, Childs B, Isaacs WB, et al. Hereditary prostate cancer: epidemiological and clinical features. J Urol 1993;150:797–802.

(26) Gronberg H, Damber L, Damber JE. Studies of genetic factors in prostate cancer in a twin population. J Urol 1994;152:1484–7.

(27) Ahlbom A, Lichtenstein P, Malmstrom H, Feychting M, Hemminki K, Pedersen NL. Cancer in twins: genetic and nongenetic familial risk factors. J Natl Cancer Inst 1997;89:287–93.

(28) Page WF, Braun NN, Partin AW, Caporaso N, Walsh P. Heredity and prostate cancer: a study of World War II twins. Prostate 1997;33:240–5.

(29) Carter BS, Ewing CM, Ward WS, Treiger BF, Aalders TW, Schalken JA, et al. Allelic loss of chromosomes 16q and 10q in human prostate cancer. Proc Natl Acad Sci U S A 1990;87:8751–5.

(30) Isaacs WB, Carter BS. Genetic changes associated with prostate cancer in humans. Cancer Surv 1991;11:15–24.

(31) Bergerheim US, Kunimi K, Collins VP, Ekman P. Deletion mapping of chromosomes 8, 10, and 16 in human prostatic carcinoma. Genes Chromosom Cancer 1991;3:215–20.

(32) MacGrogan D, Levy A, Bostwick D, Wagner M, Wells D, Bookstein R. Loss of chromosome arm 8p loci in prostate cancer: mapping by quantitative allelic imbalance. Genes Chromosom Cancer 1994;10:151–9.

(33) Baretton GB, Valina C, Vogt T, Schneiderbanger K, Diebold J, Lohrs U. Interphase cytogenetic analysis of prostatic carcinomas by use of nonisotopic *in-situ* hybridization. Cancer Res 1994;54:4472–80.

(34) Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE, and the Breast Cancer Linkage Consortium. Risks of cancer in BRCA1-mutation carriers. Lancet 1994;343:692–5.

(35) Langston AA, Stanford JL, Wicklund K, English JD, Blazej R, Ostrander EA. Germline BRCA1 mutations in selected men with prostrate cancer. Am J Hum Genet 1996;58:881–4.

(36) Berube NG, Speevak MD, Chevrette M. Suppression of tumorigenicity of human prostate cancer cells by introduction of human chromosome del (12) (q13). Cancer Res 1994;54:3077–81.

(37) Ichikawa T, Ichikawa Y, Dong J, Hawkins AL, Griffin CA, Isaacs WB, et al. Localization of metastasis suppressor gene(s) for prostatic cancer to the short arm of human chromosome 11. Cancer Res 1992;12:3486–90.

(38) Ichikawa T, Ichikawa Y, Isaacs JT. Genetic factors and suppression of metastatic ability of prostatic cancer. Cancer Res 1991;51:3788–92.

(39) Ichikawa T, Nihei N, Suzuki H, Oshimura M, Emi M, Nakamura Y, et al. Suppression of metastasis of rat prostatic cancer by introducing human chromosome 8. Cancer Res 1994:54:2299–302.

(40) Smith JR, Freije D, Carpten JD, Gronberg H, Xu J, Isaacs SD, et al. Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 1996;274:1371–4.

(41) Cannon-Albright L, Eeles RA, Neuhausen S, Goldgar DE, Lewis CM, Skolnick MH. Test of linkage between candidate loci and a prostate cancer susceptibility locus in a set of high risk kindreds. Am J Hum Genet 1994; 55:A147.

(42) Bookstein R, Bova GS, MacGrogan D, Levy A, Isaacs WB. Tumour-suppressor genes in prostatic oncogenesis: a positional approach. Br J Urol 1997;79:28–36.

(43) Schaid DJ, McDonnell SK, Blute ML, Thibodeau SN. Evidence for autosomal dominant inheritance of prostate cancer. Am J Hum Genet 1998;62: 1425–38.

(44) Gronberg H, Damber L, Damber JE, Iselius L. Segregation analysis of prostate cancer in Sweden: support for dominant inheritance. Am J Epidemiol 1997;146:552–7.

(45) Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 1996;58:1347–63.

(46) Gronberg H, Xu J, Smith JR, Carpten JD, Isaacs SD, Freije D, et al. Early age at diagnosis in families providing evidence of linkage to the hereditary prostate cancer locus (HPC1) on chromosome 1. Cancer Res 1997;57: 4707–9.

(47) Gronberg H, Isaacs SD, Smith JR, Carpten JD, Bova GS, Freije D, et al.

Characteristics of prostate cancer in families potentially linked to the hereditary prostate cancer 1 (HPC1) locus. JAMA 1997;278:1251–5.
(48) Laniado M. Prostate cancer potentially linked to the HPC1 gene [letter]. JAMA 1998;279:507–8.
(49) McIndoe RA, Stanford JL, Gibbs M, Jarvik GP, Brandzel S, Neal CL, et al. Linkage analysis in 49 high-risk families does not support a common familial prostate cancer susceptibility gene at 1q24–25. Am J Hum Genet 1997;61:347–53.
(50) Goode EL, Stanford JL, Gibbs M, Chakrabarti L, McIndoe RA, Kolb S, et al. Linkage analyses of 150 High-risk prostate cancer families at 1q24–25. Genet Epidemiol. In press 1999.
(51) Boehnke M. Estimating the power of a proposed linkage study: a practical computer simulation approach. Am J Hum Genet 1986;39:513–27.
(52) Ploughman L, Boehnke M. Estimating the power of a proposed linkage study for a complex genetic trait. Am J Hum Genet 1989;44:543–51.
(53) Cooney KA, McCarthy JD, Lange E, Huang L, Miesfeldt S, Montie JE, et al. Prostate cancer susceptibility locus on chromosome 1q: a confirmatory study. J Natl Cancer Inst 1997;89:955–9.
(54) Hsieh CL, Oakley-Girvan I, Gallagher RP, Wu AH, Kolonel LN, Teh CZ, et al. Re: prostate cancer susceptibility locus on chromosome 1q: a confirmatory study. J Natl Cancer Inst 1997;89:1893–4.
(55) Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241–7.
(56) Eeles RA, Durocher F, Edwards S, Teare D, Badzioch M, Hamoudi R, et al. Linkage analysis of chromosome 1q markers in 136 prostate cancer families. Am J Hum Genet 1998;62:653–8.
(57) Berthon P, Valeri A, Cohen-Akenine A, Drelon E, Paiss T, Wohr G, et al. Predisposing gene for early onset prostate cancer localized on chromosome 1q42.2–43. Am J Hum Genet 1998;62:1416–24.
(58) Thibodeau SN, Wang Z, Tester DJ, French AJ, Schroeder JJ, Bissonet AS, et al. Linkage analysis at the HPC1 locus in hereditary prostate cancer families. Am J Hum Genet Suppl 1997;61:A297.
(59) Xu J, Meyers D, Freije D, Isaacs S, Wiley K, Nusskern D, et al. Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 1998;20:175–9.
(60) Gibbs M, Stanford JL, McIndoe RA, Jarvik GP, Kolb S, Goode EL, et al. Evidence for a rare prostate cancer susceptibility locus at chromosome 1p36. Am J Hum Genet 1999;64:776–87.
(61) Gibbs M, Chakrabarti L, Stanford JL, Goode EL, Kolb S, Schuster EF, et al. Analysis of chromosome 1q42.2–43 in 152 families with a high risk of prostate cancer. Am J Hum Genet 1999;64:1087–95.
(62) Stanford JL, Stephenson RA, Coyle LM, Cerhan J, Correa R, Eley JW, et al. Prostate Cancer Trends 1973–1995, SEER Program National Cancer Institute. NIH Pub Bethesda (MD); 1998.

Notes

Supported by Public Health Service training grant 5T32CA09168 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services (to E. L. Goode); awards from the CaPCURE Foundation (to J. L. Stanford, L. Hood, and E. A. Ostrander); and with additional support from the Fred Hutchinson Cancer Research Center (to E. L. Goode). G. P. Jarvik is the recipient of awards from the Markey Foundation and the Pew Scholars Program.

We thank the participating PROGRESS families, whose generosity and cooperation made this investigation possible. We also thank Cassandra Neal for her expert technical expertise and Michael Brannan and Laurie Hunter for their work in accruing PROGRESS families.

Study-Design Issues in the Development of the University of Southern California Consortium's Colorectal Cancer Family Registry

Robert W. Haile, Kimberly D. Siegmund, W. James Gauderman, Duncan C. Thomas

The University of Southern California Consortium is a participating center in the National Cancer Institute's Collaborative Family Registry for Colorectal Cancer Studies (CFRCCS). Because data collection takes time, money, and effort, all of which are in short supply, we first defined our research objectives and then attempted to design our registry to enable us to address these objectives in an efficient manner. We decided on a family-based design, and our objectives are to characterize cloned genes that are generally accepted causes of colorectal cancer, to assess putative candidate genes, to map new genes, and to conduct prevention trials in high-risk subjects. For the gene characterization objectives, our primary aim is to estimate gene frequency and penetrance, with a secondary aim to investigate factors that may affect penetrance (allele-specific effects plus gene–gene and gene–environment interactions). We describe a multiple-stage design to select families into the registry. After a family is selected into the registry, we collect questionnaire data and blood samples on selected subjects only, and we tailor data collection decisions to each family (given who is affected and who is available) to optimize power per unit effort and cost. We also discuss practical decisions faced by our registry, including 1) defining a reference period for use in questionnaires; 2) deciding whether or not to establish cell lines and, if so, on whom; and 3) determining which cases should be tested for microsatellite instability. Finally, we address the appropriate use of data derived from high-risk clinics, within more broadly defined, population-based research. [Monogr Natl Cancer Inst 1999;26:89–93]

The University of Southern California (USC) Consortium is a participating center in the National Cancer Institute's Collaborative Family Registry for Colorectal Cancer Studies (CFRCCS). Other centers in the CFRCCS are the Australasian Registry, encompassing centers in Australia and New Zealand; the Fred Hutchinson Cancer Center in Seattle; the Hawaii Cancer Research Center; the University of Toronto; and the Mayo Clinic. Member institutions of the USC Consortium are the Cleveland Clinic Foundation (CCF), Dartmouth, University of Arizona, University of Colorado, University of Minnesota, University of North Carolina, and the University of Southern California. Comments here pertain to deliberations undertaken by the USC Consortium in the development of its registry and will be limited to issues of study design (another manuscript, separate from this monograph, is in preparation to discuss practical, logistical, and ethical issues addressed by our consortium).

Our consortium comprises one high-risk clinic, the CCF, and six population-based registries that cover the states of Arizona, Colorado, Minnesota, and New Hampshire, 33 counties in North Carolina, and the County of Los Angeles in Southern California. Combined, the registries ascertain more than 6000 cases of colorectal cancer per year, including approximately 500 African-American, 100 Asian, 400 Latina/o, and 5000 non-Hispanic white case patients per year. From these registries, we plan to screen 5000 case patients with colorectal cancer for a family history of colorectal cancer (50% of white case patients diagnosed at or above 50 years of age in 1996, plus 100% of case patients diagnosed under 50 years of age and all African-American, Asian, and Latina/o case patients over a 4-year period, from 1996 through 1999. The CCF has 382 families with familial adenomatous polyposis (3250 individuals), 100 families with hereditary nonpolyposis colorectal cancer (HNPCC) (900 subjects), and 250 families with "HNPCC-like"/non-Amsterdam criteria (2250 subjects).

Because data collection takes time, money, and effort, all of which are in limited supply, we first defined our research objectives and then attempted to design our registry to enable us to address these objectives in a very efficient manner. Broadly defined, our research objectives are to characterize cloned genes that are generally accepted causes of colorectal cancer, such as the mismatch repair genes, to assess putative candidate genes, to map new genes that cause colorectal cancer, and to conduct prevention trials in high-risk subjects. For accepted and putative cancer genes, our primary aims are to estimate gene frequency and penetrance and to investigate factors that affect penetrance (allele-specific effects plus gene–gene and gene–environment interactions).

Design: Population-Based Registries

Aims

With these objectives in mind, Drs. Thomas, Gauderman, and Siegmund at USC entered discussions and analyses with Dr. Haile to determine an optimal design for the registry. For reasons described in another paper *(1)*, we favored the use of a family-based design to meet our objectives (no unrelated, population-based control subjects). We were aware of another design option that includes control families but did not consider it further because it was our intuition that such control families do not provide enough additional information to warrant the increased costs (in terms of fieldwork and laboratory work) of including them, but this belief merits further analysis and consideration.

Affiliation of authors: Department of Preventive Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles.

Correspondence to: Robert W. Haile, Ph.D., Department of Preventive Medicine, University of Southern California/Norris Cancer Center, 1441 Eastlake Ave., MS#44, Rm. 4455, Los Angeles, CA 90033 (e-mail: haile@hsc.usc.edu).

© Oxford University Press

Within the general design framework (a family-based study in which the families are ascertained through a case patient), the researchers addressed questions of what proportions of single-case versus multiple-case families to include in the registry and, within families, on which subjects to collect risk-factor questionnaires, blood samples, tumor blocks, and other data.

Detailed results of deliberations are presented in an accompanying paper by Siegmund et al. *(2)*. We discuss here some of the actual issues we dealt with in applying a multistage design to a "real life" circumstance, with reality-based constraints and practical considerations. The first question addressed was what types of families (defined by family history of colorectal cancer) should be sampled in what proportions to achieve an optimal design. In general, if one were primarily interested in estimating gene frequency or penetrance accurately, results suggest including a substantial number of single-case families because the variance of the estimator is greater in these families (a lower proportion carries the high-risk allele or mutation). In contrast, if one were primarily interested in gene–covariate interactions or mapping new genes, one would want to start with families with the most case patients and then ascertain families with fewer and fewer case patients, as the "heavily loaded" families are exhausted. It was clear that no single sampling strategy was optimum for all objectives. At this point, the principal investigators of each center in the consortium, along with Dr. Thomas, had to prioritize the research objectives. There was a strong consensus that obtaining accurate estimates of gene frequency and penetrance should be our highest priority (it should be noted that all P.I.s are epidemiologists and take a population-based perspective of this research). There was also a strong interest in retaining a reasonable ability to investigate gene–environment interactions. Mapping of new genes was considered the lowest priority for this registry because we believed other efforts were in a better position to find new genes and that we should concentrate on our strength of population-based research on cloned genes. With these priorities in mind, a multistage sampling scheme was recommended, as described in detail by Siegmund et al. *(2)*. We comment on relevant issues here.

Multistage Sampling Protocol

The first step in the registry is to screen 5000 case patients with colorectal cancer for a family history of cancer. Given the large number of case patients to screen, it was imperative that we design a screening questionnaire that was quick and relatively accurate, understanding that we could not afford to verify all reported cancers at this stage of the study. To keep the questionnaire short, we decided to ask probands only about their siblings and parents and to ask only about colorectal cancer. For purposes of selecting families for our second-stage sample, any classification strategy that was applied equally to all case patients would not create any biases. Therefore, we could safely ignore other cancers and cancer history in second-degree relatives. It was also our feeling, not based on formal simulations, that including children of probands at this stage was not worth the effort, given the expected low yield of case patients. We also decided not to ask about polyps because information on polyp status is usually highly selected by family history, access to care, and other variables that are difficult to quantify. A key point here is that it does not matter if the information obtained at this stage is "accurate" in the multistage sampling paradigm suggested for our registry, provided one uses the measured, not the "true," family history information to compute the sampling fractions as weights in the analysis. A sensitivity analysis of our design showed that underreporting and overreporting of family history (7% and 27%, respectively) resulted in only a slight loss in statistical efficiency that was partially overcome by a larger sample size and increased total costs (Siegmund KD: unpublished data). Therefore, even if having more accurate information initially would lead to a statistically more efficient sampling strategy, these inaccuracies would not produce any bias if analyzed correctly.

The screening survey will identify with some error nuclear families with one, two, three, or more cases of colorectal cancer. As described by Siegmund et al. *(2)*, we decided to employ a sampling scheme that would have us include 16% of single-case, 32% of two-case, 48% of three-case families, and so on, for further data collection. This sampling scheme would yield a set of families that provides reasonable power to estimate gene frequency and penetrance [*see* Table 4, *(2)*] for a reasonable range of these parameters and also adequate power to estimate gene–environment interactions for reasonable ranges of effects, particularly when analyses can include case-cousin pairs in addition to case-sib pairs and case-sib triplets.

Given that we now had a sampling scheme for families, the next set of questions was on which subjects within families to collect questionnaire data and blood samples. Again, this selection procedure is described in detail by Siegmund et al. *(2)*. We were guided by the principle that we should tailor data-collection decisions to each family (given who is affected and who is available) to optimize power for unit effort and cost. As one example, if an unaffected sibling were available, we would want to obtain both a blood sample and a questionnaire from that subject, who would serve as a control in case–control comparisons. If no such control were available, but parents of the proband were available, we would obtain a blood sample from them as they are informative for some genetic analyses, but we would not obtain risk-factor questionnaires because they would not be comparable to the case patients' questionnaire, given the generational difference and secular trends in selected exposures of interest. For example, assume a case patient was diagnosed in 1999 at 40 years of age and she had a mother who was 70 years. Asking about history of oral contraceptive use during her 20s would be problematic because oral contraceptives were not widely available when the mother was in her 20s (the exposure prevalence in the source population of the case patient would be much higher than in the source population of the mother) *(3–5)*. (We note here that gene–environment interactions may, in principle, be examined in parent–child trios with exposure data only on the proband, not the parents. This requires the assumption that the gene and environmental factor of interest are independently distributed conditional on the parents' genotypes, or equivalently within families, not across families as in the case-only design.) In addition to sibling controls, we would consider the use of cousin controls for estimating main effects and gene–environment interactions. This theme of tailoring our data collection to maintain validity and enhance efficiency has guided many of our decisions, as described below.

Practical Decisions

We have addressed other study-design issues that have grown out of the basic decisions described above. We briefly address

three here to illustrate the types of decisions one may face in developing a registry: 1) defining the reference period in questionnaires when asking about exposures or practices at one point in time (such as diet) as opposed to collecting a "lifetime" exposure history; 2) deciding whether or not cell lines should be established and, if so, on whom; and 3) determining which cases should be tested for microsatellite instability (MSI) or replication error (RER) phenotype.

Reference Period

Defining an appropriate reference period was problematic. A reference period is necessary to set an upper limit for obtaining information on "lifetime" exposures up to the reference period (e.g., before the diagnosis of cancer in a case patient), and it is also necessary when we are obtaining a report of behaviors or exposures that refer to one period of time (e.g., one's typical diet the year before diagnosis). Ideally, we want to set a reference period that is comparable for case patients and control subjects, that minimizes bias, and that does not compromise accuracy of recall. One distinction we face in defining "time" of exposure is between calendar time (a concern if there are secular trends in exposure or incidence) and "biologic" time, typically measured by age (a concern if there are critical ages of exposure and because we want case patients and control subjects to have similar periods at risk). An additional factor in family studies is that diagnosis of one case patient in a family may possibly cause other relatives in that family to alter their exposures if they believe those exposures may affect risk of disease (e.g., changing one's smoking behavior or diet in response to a family member's diagnosis of lung cancer or colon cancer). This family history may also influence physicians to alter advice they give to that family (e.g., possibly suggesting the use of oral contraceptives to a woman with a family history of ovarian cancer in the belief it will reduce her risk of ovarian cancer). In our opinion, there is no simple and completely correct definition of the reference period. For "lifetime" exposures, one may collect "complete" data up to the present and then treat the calculation of appropriate individual-specific reference dates as an analysis issue. For some exposures, such as diet, this option is not feasible.

The option recommended by the Epidemiology Working Group for all centers of the CFRCCS was to use "2 years ago" as the reference period for all subjects. The arguments for doing this were that it was a simple standard for all subjects (cases and controls) and avoided referring to the date of diagnosis, which some investigators believe would reduce the potential for recall bias. There are arguments against it.

One argument is that it is naive to believe that the case patients are not using their date of diagnosis as a reference period in their minds because it is such an important event to them. Therefore, we ought to use the date of diagnosis because it provides a more natural reference period than "2 years ago" that has an arbitrary quality about it, especially for case patients. This arbitrary quality may make it difficult for all subjects to relate to it, thereby compromising the accuracy of responses. A second argument is that, within families, one subject may complete a questionnaire in the current year, while another subject, for a host of reasons (divorce, ill health, and changing jobs are some reasons we have encountered), may delay completing the questionnaire for a number of years; therefore, the reference period now refers to a different calendar time.

An alternative is to define the reference period as the year before the diagnosis for case patients. Some investigators worry about recall bias because this reference period means more to a case patient than to a control subject. Also, within multiple-case families, case patients may have been diagnosed many years apart, even within sibships, so case patients are left referring to different periods of time (either calendar time or age, depending on how the reference period is defined). Even if case patients are asked to refer to a period of time before diagnosis, the situation of what to do with control subjects is unresolved. The most natural period for them is to recall events, such as diet, within the past year.

Some researchers worry about the lack of comparability with case patients if they are referring to a period a few years ago, before diagnosis, and control subjects are referring to the current time, especially for exposures with strong secular trends. Setting the reference period for control subjects based on the date of diagnosis for the case patient in the family is another option. Here, we may be asking control subjects about exposures for a period of time that may seem arbitrary and removed from the present.

As we stated at the beginning, no simple solution is available that satisfies all investigators. The CFRCCS has decided to use the reference period of 2 years ago, with the option of stretching this reference period to 3 years to avoid asking case patients about exposures after their diagnosis.

Cell Lines

Another question that we face is on which subjects to establish cell lines that offer, in principle, an infinite source of DNA and RNA. Because characterizing cloned genes, as opposed to finding new ones, is a higher priority, we will limit comments to research with cloned genes. One purpose of the registry is to evaluate possible candidate genes, with association studies of main effects and possibly gene–gene and gene–environment interactions. We usually use a polymerase chain reaction-based technique for genotyping the polymorphism of interest, so we would probably not run out of DNA (provided we collect enough in the beginning) for this purpose; therefore, establishing and maintaining cell lines would probably not be justified for this type of research.

Another major purpose of the registry is to characterize, at the population level, relatively rare susceptibility genes that have already been identified as causes of colorectal cancer (such as the mismatch repair genes) in terms of gene frequency, penetrance, and gene–covariate interactions. With this purpose in mind, we face the prospect of screening large genes for relatively rare mutations. This process can use up a lot of DNA to identify and confirm the nature of the sequence variation. It is for this type of research that cell lines may be justified. We see two extreme options and various scenarios in between the extremes. One extreme is to establish a cell line only on the proband in the family. If we actually ran out of DNA on other family members but had it on the proband, we could still provide model-based estimates of gene frequency and penetrance. One could also explore gene–environment interactions if we were prepared to assume gene–environment independence within families, as discussed above. At the other extreme, we could establish cell lines for "all" family members who would contribute useful information to our planned analyses (frequency, penetrance, and interactions). The

obvious disadvantages are the substantially increased labor and costs.

At least two less extreme options are worth considering. One is to establish cell lines only for the proband and require the initial search for mutations to use DNA (or RNA) only from the proband. Once a mutation is identified, more efficient assays can be used to test for that mutation in all other family members. Valid case–control comparisons would, however, have to be restricted to mutations that could be detected using the same methodology for cases and controls. The other option is to tailor the establishment of cell lines on a family-by-family basis, depending on the structure of the family and the intended analyses, which is similar to what we are planning to do for collection of blood samples and questionnaires.

Tumor Blocks

Another issue that we are addressing is the collection of tumor blocks and the testing for MSI, also referred to as RER. This activity becomes an issue because the collection and processing of tumor blocks takes time and money. A consensus at the CFRCCS is that we should determine MSI status for all cases (when isolated cases are entered into the registry) or for all families in the CFRCCS. If germline mutations in a mismatch repair gene were the only interest, perhaps collecting tumor blocks from one case patient with colorectal cancer per family and testing for MSI would suffice as a relatively accurate classification of that family with respect to MSI status (for purposes of detecting germline mutations that are responsible for the MSI). This approach would not work, however, for classifying cases in which the MSI is caused by alternative (nongermline) events, such as hypermethylation of the promoter region of hMLH1, that may or may not track in families. Collecting tumor blocks from other case patients with colorectal cancer (in addition to the proband) would enable us to investigate the prevalence of MSI (by all causes), sharing of MSI status among cases, as well as sharing of other molecular events, which could help elucidate modifier genes and pathogenetic mechanisms associated with colorectal cancers with MSI. Collecting tumor blocks from other sites of cancer would enable us to extend the investigations to these cancers as well, shedding light on common mechanisms and helping us to define what cancers are actually associated with HNPCC and other syndromes. Currently, our consortium has decided to try to obtain tumor blocks for all reported cancers in family members we deem potentially informative for planned analyses, but we are tracking the costs and may amend this decision in the future.

In summary, with respect to our population-based research, we started by defining our objectives. The reality that we cannot address all objectives equally well forced us to set priorities. Within these priorities, we tailor decisions about sampling and data collection to the family structure (which relatives are affected and which ones are likely to be available for data collection) in an attempt to maximize our power for planned analyses per unit cost and effort while retaining a valid design.

High-Risk Clinic

A final, major issue we are addressing is the appropriate use of data derived from high-risk clinics, such as those available from the CCF, a member institution in our consortium. In our opinion, such data may be quite useful for multiple purposes. First, research focused and applied to high-risk families is important, per se, because we need more information on how to best serve members of these families. For example, our consortium is interested in prevention trials specifically in high-risk subjects. Two means of defining high risk are a positive family history or actual genetic data that identify a subject as carrying a mutation in a gene of interest. High-risk registries generally provide a higher yield of subjects with a positive family history and carriers of mutations than a sample from the general population. Results from such trials would be most directly applicable to other, similarly defined subjects.

A broader question is whether results obtained from studies of subjects identified through high-risk clinics (with ascertainment schemes characterized at this meeting as "indescribable") can be used to estimate meaningful population parameters. We start by considering the generalizability of results from prevention trials. Here the main question would be the prevalence of effect modifiers and whether the prevalence of actual modifiers differs to an important degree between the high-risk families and other subjects to whom one may want to generalize the results. The problem is that we often do not know what these modifiers are and, therefore, have to make judgments about generalizability without this knowledge.

Another question is whether we can derive valid estimates of gene frequency and penetrance from these types of families. In principle, one may be able to do so if the correct analysis is performed and the model underlying the analysis is correctly specified. It may be instructive here to consider the results published to date about the estimated penetrance of BRCA1. In the analysis by Easton et al. *(6)*, penetrance was estimated to be 82% by age 70 years, based primarily on data from highly selected high-risk families. In the analysis of BRCA1 mutations in Ashkenazi Jewish subjects by Streuwing et al. *(7)*, penetrance was estimated to be about 56% by age 70 years. Results from a population-based study in Australia, which is part of the Collaborative Family Registry for Breast Cancer Studies, suggest that the penetrance of protein-truncating mutations is around 40% *(8)*. Because penetrance drives so many decisions about how to manage carriers, it is important to resolve or to understand the basis of these different estimates.

It seems that at least two reasons are likely for the estimate of 85% by Easton et al. *(6)*. It is possible that there is heterogeneity in the penetrance of BRCA1 mutations, defined by modifying factors, either genetic or environmental. High-risk families may be selected for modifying factors that increase penetrance. In this case, the estimates by Easton et al. may be correct for the types of families included in this analysis. The estimates by Streuwing et al. *(7)* may be lower because the specific mutations studied, which are more common in Ashkenazi Jews, may be less penetrant than other mutations or because the particular Ashkenazi Jewish populations studied may have other protective modifying factors. It is also conceivable that the analyses failed to adequately take into account residual familial correlations, which could bias penetrance upward. These two explanations are not mutually exclusive. In fact, it seems plausible that there may be some important modifying factors and that these factors may correlate within families. Given the experience with BRCA1 to date, it would seem prudent not to rely on estimates of gene frequency and penetrance derived from variously ascertained high-risk families until we better understand determinants of penetrance and how these determinants are associated with ascertainment.

This brings us to a consideration of determinants of penetrance. Can findings regarding gene–gene and gene–environment interactions derived from variously ascertained high-risk families be generalized to some broader population? Here again, we must consider the question of modifiers, in this case, modifiers of the interaction. We must ask how strong are these possible modifiers and how likely are they to be differentially distributed in a substantial manner. Because we are often operating in ignorance of the answers to these questions, it would seem unwise to generalize results. In our consortium, we plan to investigate gene–covariate interactions in data from the CCF (because the yield of carriers is expected to be higher than a population-based series of families) as a source of leads that may be followed up in our population-based families. In this manner, we can begin to build an empirical basis for seeing what interaction effects are similar or different between variously ascertained high-risk families and population-based series of families. In summary, high-risk families derived primarily from clinical settings are certainly a prevalent and potentially valuable resource, so we need careful thought about their proper use and integration into the population-based research many of us are conducting.

References

(1) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Monogr Natl Cancer Inst 1999;26:31–7.

(2) Siegmund KD, Whittemore AS, Thomas DC. Multistage sampling for disease family registries. Monogr Natl Cancer Inst 1999;26:43–8.

(3) Piper JM, Kennedy DL. Oral contraceptives in the United States: trends in content and potency. Int J Epidemiol 1987;16:215–21.

(4) Ursin G, Aragaki CC, Paganini-Hill A, Siemiatycki J, Thompson WD, Haile RW. Oral contraceptives and premenopausal bilateral breast cancer: a case–control study. Epidemiology 1992;3:414–9.

(5) Ursin G, Bernstein L, Pike MC. Breast cancer. Cancer Surveys. Trends in Cancer Incidence and Mortality 1994;19/20:241–64.

(6) Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BCA1 mutation carriers. Am J Hum Genet 1995;56:267–71.

(7) Streuwing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, et al. Cancer risk with 185delAG and 5382insC mutations of BRCA1 and the 6174delT mutation of BRCA2 among Ashkenazi Jews. N Engl J Med 1997;336:1401–8.

(8) Hopper JL, Southey MC, Dite GS, Jolley DJ, Giles GG, McCredie MR, et al. Population-based estimate of the average age-specific cumulative risk of breast cancer for a defined set of protein-turncating mutations in BRCA1 and BRCA2. Australian Breast Cancer Family Study. Cancer Epidemiol Biomarkers Prev 1999;8:741–7.

Design and Analysis Issues in a Population-Based, Case–Control–Family Study of the Genetic Epidemiology of Breast Cancer and the Co-operative Family Registry for Breast Cancer Studies (CFRBCS)

John L. Hopper, Georgia Chenevix-Trench, Damien J. Jolley, Gillian S. Dite, Mark A. Jenkins, Deon J. Venter, Margaret R. E. McCredie, Graham G. Giles

Background: **Historically, studies of the "genetic epidemiology" of cancer have used nonsystematically sampled kindreds with numerous cases of cancer across multiple generations. From the epidemiologic viewpoint, it is difficult to extrapolate findings to the population because of the *ad hoc* ascertainment of these atypical, ill-defined families. Since 1992, we have been conducting a population-based, case–control–family study of breast cancer.** ***Methods:*** **Families are identified through a single, population-sampled proband, who is either affected or unaffected, making adjustment for ascertainment straightforward. Administered questionnaires and blood samples are sought from cases, controls, and specified sets of relatives. From 1996 through 1999, a further 1200 case families have been recruited as part of the Co-operative Family Registry for Breast Cancer Studies (CFRBCS). Issues relevant to the study design and analysis are discussed.** ***Results:*** **Epidemiologic and genetic findings published to date are summarized. In particular, this population-based study has shown that the so-called "high-risk" families containing multiple cases of breast cancer are *not* typical of families in the general population in which BRCA1 or BRCA2 mutations are segregating. Most "hereditary" cancers are "sporadic."** ***Conclusion:*** **The collection of DNA, as well as data on disease status and risk factors, from population-sampled sets of relatives provides a powerful resource for addressing genetic and environmental determinants of cancer. A population-based multicenter, multidisciplinary enterprise, such as has been developed by the CFRBCS, may become a model for future research in cancer epidemiology, allowing genetic and environmental risk factors to be put into a proper population perspective. [Monogr Natl Cancer Inst 1999;26:95–100]**

In 1982, Morton *(1)* defined the new discipline of "genetic epidemiology" by modifying the definition of "epidemiology" given by Last *(2)*. It was to be "a science that deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations." In this context, "inherited" was meant to include biologic and nonbiologic inheritance *(1)*. That is, in genetic epidemiology the basic unit of study is the family (as distinct from the individual), and both genetic and nongenetic causes of familial aggregation are to be considered. Clearly, the method of sampling families (i.e., ascertainment) is a key issue.

Over the past few decades, however, studies purporting to be of the genetic epidemiology of cancer have almost exclusively used families in which there are a number of cases of one or more types of cancer, spread across multiple generations. This strategy has been important, as it has led to the identification of "familial cancer syndromes" and of some of the genes causing cancer in a proportion of these atypical kindreds. From an epidemiologic viewpoint, it is difficult to extrapolate findings from these families—such as estimates of disease risk in mutation carriers—to any defined set of people, let alone the general population. This is because of the lack of a clear or consistent definition of sampled families and because of the *ad hoc* (i.e., nonrandom) ascertainment, violating the basic principles of statistical inference. Whereas these genetic studies have been important, it is difficult to classify them as part of "genetic epidemiology," where "epidemiology" is a noun, and "genetic" is an adjective meant to also include consideration of nongenetic factors *(1)*.

To be able to make valid conclusions about the roles of genetic factors in regard to breast cancer in the population and to put the roles of genetic and nongenetic factors into their proper perspective, we have been conducting a population-based study of the genetic epidemiology of breast cancer *(3,4)*. In this paper, we will describe the study and discuss some design and analysis issues based on our experiences.

Australian Breast Cancer Family Study (ABCFS)

Families are identified through a single, population-sampled proband, who is either affected or unaffected. This sampling makes adjustment for ascertainment straightforward, simply by conditioning on the proband's age and disease status. Case probands are women recently diagnosed with a first primary, invasive breast cancer reported to the Victorian or New South Wales Cancer Registry and living in Melbourne or Sydney, respectively, at the time of their diagnosis. Cancer registration is compulsory under legislation and is considered to be complete. Con-

Affiliation of authors: J. L. Hopper, D. J. Jolley, G. S. Dite, M. A. Jenkins, Centre for Genetic Epidemiology, The University of Melbourne, Carlton, Australia; G. Chenevix-Trench, Queensland Institute of Medical Research, and Department of Pathology, University of Queensland, Herston, Australia; D. J. Venter, Department of Pathology and Research, Peter MacCallum Cancer Institute, Melbourne, Australia; M. R. E. McCredie, New South Wales Cancer Council, Woolloomooloo, Australia, Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand; G. G. Giles, Cancer Epidemiology Centre, The Anti-Cancer Council of Victoria, Carlton, Australia.

Correspondence to: John L. Hopper, Ph.D., The University of Melbourne, Centre for Genetic Epidemiology, 200 Berkeley St., Carlton, Victoria 3053, Australia (e-mail: j.hopper@gpph.unimelb.edu.au).

© Oxford University Press

trol probands are women without breast cancer, selected at random from the electoral rolls for the relevant cities (adult registration for voting is compulsory in Australia). Questionnaires are used to measure epidemiologic risk factors in case probands, control probands, and relatives of *both* types of probands. We call this a "case–control–family" design *(3)*. Blood samples are collected from case probands and control probands and from specified sets of relatives depending on the family cancer history. In this design, strict rules define which relatives are to be studied (*see below*).

Why Did We Do the Study?

At the time the ABCFS commenced (initial grants were written in the late 1980s, and pilot work and recruitment began in 1992), the expression "BRCA1" had yet to be coined. Our primary aims at the time were to collect the best possible epidemiologically valid data to study the effect of family history on risk of breast cancer and to obtain a resource of DNA from population-based families for future research once genes associated with breast cancer susceptibility were identified. Previously, the typical epidemiologic approach had been to rely solely on unverified self-reports collected as part of conventional case–control and cohort studies in which family history was only one of a large number of putative risk factors being addressed.

We also wished to apply the new "regressive logistic" models for analysis of family data *(5,6)* to test empirically the extent to which familial aggregation in risk factors might account for familial aggregation of cancer. [A paper on the theoretical aspects of this issue was published in 1992 *(7)*]. These new statistical methods, based on logistic regression and fitted by maximum likelihood, also allow determination of the extent to which the *residual* familial aggregation can be attributed to an *unmeasured* genetic effect (commonly referred to as a "major gene effect"). The residual familial aggregation is that remaining after adjusting for risk factors *measured* on individual family members, with or without adjustment for the effect of parental disease status on an individual's risk, and for disease dependencies within sibships *(5)*. The risk factors could be measured genetic variants, measured environmental or lifestyle factors, or a combination of both. For example, if mutation status for the currently identified breast cancer susceptibility genes, BRCA1 and BRCA2, is known for case probands and at least some of their relatives, this design and approach to statistical analysis will allow us to address questions such as how much familial aggregation is explained by BRCA1 and BRCA2 mutations; what evidence is there for other "major breast cancer genes"; and what is the most likely penetrance and prevalence of mutations in these, as yet, unknown genes?

To optimize information on genetic factors, we started by sampling women with early onset breast cancer, i.e., breast cancer diagnosed before the age of 40 years. Case–control and cohort studies have consistently indicated that the multiplicative increase in risk to close female relatives of an affected woman is greater for early onset disease, and the earlier the onset of disease in the affected relative. Furthermore, segregation analyses have suggested that the proportion of breast cancer attributable to a major gene is greater the earlier the onset of disease. In recent years, we have been able to extend the sampling to case probands with later age at diagnosis (*see below*).

Why Did We Not Use a Two-Stage Design, Based on Oversampling Women With Breast Cancer Who Self-Report Having a Family History?

Although oversampling affected individuals with a family history of the disease will be more efficient for studying familial factors *(8,9)*, in practice this strategy will become problematic if the decision to sample is based on incorrect information. Previous experience in Australia with a population-based family study of malignant melanoma found that, for a high percentage of respondents, their knowledge of malignant melanoma in relatives could not be relied on; the false-positive reporting of family history by affected case probands was 40% *(10)*. Although we did not expect the problem of relying on reports of disease in relatives to be of the same magnitude for breast cancer, it served as a warning, particularly as we wanted to extend our pedigrees to more than first-degree relatives of probands. In practice, studying all cases irrespective of family history has made conduct and analysis easier, but it may have led to inefficiencies in data collection for some analyses. However, given that the study was being developed as a resource to address a number of hypotheses, some of which would involve genes and genetic markers not yet discovered, the extra cost and effort may in the long term prove to be worthwhile.

Why Study Population-Based Control Families?

Epidemiology is built around the comparison of affected and unaffected subjects. Therefore, we considered that, as we were studying relatives of affecteds (case families), we should also study relatives of unaffecteds (control families). Furthermore, and according to the principles of modern epidemiology, case families and control families should be studied using as far as practical identical methods. This approach was motivated by our principal concerns: to measure the population cancer burden associated with genes; to study concurrently the roles and possible interactions of both genetic and environmental risk factors on cancer susceptibility; and to identify genetic and environmental factors and, in particular, modifiable lifestyle factors that ameliorate or exacerbate susceptibility to cancer in women at high genetic risk. Such information is necessary to determine the relative costs and benefits of different genetic screening programs and to help plan optimal strategies.

Whom to Interview and From Whom to Take a Blood Sample?

A number of design issues were addressed during the conduct of the study. Major questions revolved around how wide to extend the core family (i.e., whether or not to include aunts, uncles, grandparents, nephews, and nieces) and who to take blood from and who to interview.

In pilot work, we studied the case proband and her sisters, brothers, parents, and grandparents. It was soon realized that, given the young age at diagnosis, only a small proportion had an affected female relative (<15%). Therefore, although it meant a substantial increase in the number of interviews per family, it was decided to extend the sampling frame to include aunts.

Table 1 shows, for each type of relative, the proportions for whom a personal interview or a proxy interview was conducted and the proportion deceased, for the relatives of the 248 women with breast cancer diagnosed before age 40 years, studied between 1992 and 1995 and living in Melbourne. For most of this period, all sisters of cases were asked to provide a blood sample,

Table 1. Proportions of relatives of cases diagnosed before the age of 40 years, by vital status and interview status, for the Melbourne component of the Australian Breast Cancer Family Study, 1992–1995

Relative	Alive, %		Deceased, proxy, %	No data, %
	Interviewed	Proxy		
Sisters	76	22	1	1
Brothers	60	35	3	2
Mothers	65	21	14	—
Fathers	41	26	33	—
Aunts	39	36	19	4
Grandmothers	3	9	85	3
Grandfathers	1	3	92	4

and this was achieved for 60% of identified sisters. For each case, there were on average 1.4 sisters, 1.4 brothers, and 2.8 aunts.

Proxy interviews were used for persons who were deceased or who refused to participate so as to collect information on important risk factors, such as parity, that could reasonably be asked of close relatives. Male relatives were interviewed because they may be at an increased risk of breast or other cancers (at the start of the study there was little epidemiologic information to support such a conjecture, but the increased risk of breast and prostate cancers now apparently evident in male carriers of mutations in BRCA2, and BRCA1 and BRCA2, respectively, supports this decision). The male questionnaire is the same as for females, excluding questions relevant only to females.

It can be seen from Table 1 that the proportions interviewed were highest for mothers and sisters, lower for males and for the older generations of relatives, being lowest for grandparents for whom the great majority were deceased. These interview proportions are likely to be lower in the relatives of older case probands.

Of the total of 644 eligible case probands from Melbourne and Sydney studied between 1992 and 1995 *(4)*, 467 were interviewed (72.5%). Attrition was a result of death (1.7%), refusal (surgeon 8.4%; patient 11.8%), nonresponse (surgeon 0.6%; patient 1.4%), or having moved residence and not being located (3.6%). Of the 632 eligible control probands, refusals (25.8%) and nonresponse (9.8%) resulted in 408 being interviewed (64.4%). Blood samples were available from 393 case probands (84.2% of participating and 61.0% of those eligible) and 294 control probands (72.1% of participating and 46.5% of those eligible) [*see (11)*].

The final protocol for whom to interview and from whom to obtain a blood sample was as follows: In the first instance, blood is sought from both the case and control probands as well as from the mother, father, and any sisters of the case probands. (Blood is sought only from relatives of control probands if they have two or more affected relatives.) Interviews are sought from all adult first-degree relatives and from aunts and grandparents of either sex on both sides of the family. In case families, should a first- or second-degree relative be reported to have been affected with breast or ovarian cancer, blood and an interview are sought from the affected relative and all first-degree relatives from whom blood has not already been sought. Should any of these newly approached relatives be affected, blood and interviews are sought from their yet-to-be approached first-degree relatives, and so on. This sampling scheme is a valid version of the Cannings–Thompson Sequential Ascertainment Scheme in which adjustment for ascertainment is achieved by conditioning on the disease status of the proband *(12)*.

Co-operative Family Registry for Breast Cancer Studies (CFRBCS)

Since 1996, we have studied a further 1200 case families as part of the National Institutes of Health-funded CFRBCS, an international resource for studies of the genetic epidemiology of breast cancer. This resource consists of two other centers focusing on population-based studies similar to the one described above (one in San Francisco, CA, and the other in Ontario, Canada) and three U.S. centers recruiting families with extensive histories of breast cancer through family cancer clinics. Core questionnaires and uniform protocols for the collection and storage of blood and tissue have been developed. Pathology review is under way, so that issues concerning the comparability of diagnoses of cancer subtypes within and between centers and countries can be addressed. Pathology has become a more important issue following our population-based confirmation of distinct histologic phenotypes *(13)* and molecular pathogenic pathways *(14)* in breast carcinomas occurring in women with a germline mutation in BRCA1 or BRCA2.

Some Further Design Issues

A number of design issues have arisen out of the first few years' experience in establishing and conducting the CFRBCS, a multicenter project in genetic epidemiology. One important issue revolves around the need, at least for some exposures, to ask probands about their experience up until a pre-designated time. For case probands, we chose "1 year before diagnosis" as their reference age. For control probands, we have asked about their experiences up until the time of interview. In an unmatched case–control study of probands, we made the reference age for control probands their age at interview and adjusted for reference age in the analysis *(4)*.

The reference age issue is also critical when case probands are compared with unaffected sisters in a matched analysis. This is because it is possible that the sibling controls may have changed their behavior(s) following the diagnosis of the proband, or there may be secular trends in the exposure (as was historically the case for oral contraceptive use). Therefore, it will be necessary to truncate information on the exposure of sister(s) to the calendar date of the case proband's diagnosis. If, following the principles of risk-set sampling, only unaffected sisters who have attained the reference age of the case proband are eligible, they would need to have done so before the calendar time of diagnosis. Analyses can then be conducted with the use of truncated or nontruncated exposures to assess to what extent the effect of learning of one's sister's diagnosis might influence the results. These and a number of similarly important issues in using unaffected sibling or family controls are discussed in a companion paper in this monograph *(15)*.

The issue of how to word questions must also recognize that, although the study has a case–control component, it is also a family study and has the potential of becoming a cohort study of families in the long term. Therefore, questionnaire design must also take into account the likelihood of follow-up questionnaires being administered at regular intervals and updating exposures in both affected and unaffected individuals as well as the vital and disease status of all relatives. Items that need to be considered include who to follow-up and the method(s) of follow-up

(e.g., active follow-up using mailed questionnaires, telephone interviews, etc., or passive follow-up relying on identification of cancer and deaths with the use of population registries, etc.). Note that, when family history data are updated, this must be done systematically to protect from any bias that might arise from more attention being given to those families in which one is told that new cases have occurred.

Statistical Estimation of Cumulative Risk

Estimation of cumulative risk by analyzing cancer histories in a defined set of relatives of mutation carriers in which, as in our study, the mutation status of all relatives is not necessarily known poses an interesting statistical challenge. We have used two approaches: one based on repeated sampling and Kaplan–Meier curves *(16)* and the other based on maximum likelihood estimation assuming a functional form for the cumulative risk *(17)*. These methods were used to derive the penetrance estimate from analysis of time to diagnosis of breast cancer in known and potential mutation-carrying relatives and were shown to give almost identical estimates and confidence intervals *(16)*.

The likelihood approach can also be applied to estimate the risk of disease associated with other genes, even those for which the increased risk may be modest. This approach is made possible because of our defined population-based ascertainment scheme for which correct adjustment can be made, the use of population incidence rates derived by the registries from which case probands have been sampled, and the quality of the data on disease status that derives from our seeking to interview living relatives and attempting to verify all reports of cancers in the family *(4)*. Furthermore, the mode of inheritance can be specified as dominant, recessive, or codominant. The estimation of risk in terms of the hazard ratio permits interpretation in the usual case–control manner, allowing direct estimation of the attributable risk associated with genetic variants within each age category. It is not necessary to measure genetic variants for all—or even any—relatives, and there is no bias from missing genetic information, provided there is no nonpaternity or nonmaternity, because the genetic status of unmeasured relatives is inferred probabilistically, conditional on the measured genotype of the proband and that of any relative. The likelihood approach can, at least in theory, be extended to include other causes of familial aggregation and to incorporate risk factors measured on relatives, as discussed above in reference to regressive logistic models.

Results

Some Findings to Date: Epidemiology

Between 1992 and 1995, the ABCFS studied 467 women with breast cancer and 408 control probands. A standard case–control analysis *(4)* revealed that the main risk factor was having a first-degree relative with breast cancer; *see also* Table 2 in which the categories have been expanded and one case proband since found to be ineligible has been deleted. The great majority of affected first-degree relatives were mothers, and this risk factor was associated with an odds ratio of about 3. Having a sister affected was associated with a greater odds ratio, but, because of the paucity of affected sisters of control probands, the estimate had a wide confidence interval. The increased risk associated with at least one affected second-degree relative was about 1.3 but not nominally significant, although only effects greater than twofold could be excluded with reasonable statistical power.

Table 2. Estimates of familial risks of breast cancer in women before the age of 40 years, based on reported history of breast cancer in female relatives from the Australian Breast Cancer Family Study, 1992–1995

Affected relative(s)	% (No.) affected: Case subjects	% (No.) affected: Control subjects	Relative risk* (95% confidence interval)
Mother	10.7 (50)	4.9 (20)	2.9 (1.6–5.2)
Sister	3.2 (15)	0.3 (1)	16.4 (2.1–128)
First-degree	12.5 (58)	5.2 (21)	3.4 (1.9–5.9)
Maternal aunt	5.6 (26)	6.9 (28)	0.9 (0.5–1.6)
Paternal aunt	7.5 (35)	4.2 (17)	1.8 (1.0–3.4)
Any aunt	12.2 (57)	11.0 (45)	1.2 (0.8–1.9)
Maternal grandmother	6.4 (30)	5.6 (23)	1.2 (0.7–2.2)
Paternal grandmother	4.7 (22)	2.9 (12)	1.6 (0.7–3.4)
Any grandmother	10.9 (51)	8.3 (34)	1.4 (0.9–2.3)
Second-degree	21.7 (101)	18.6 (76)	1.3 (0.9–1.8)
Second-degree only	17.8 (83)	17.9 (73)	1.0 (0.7–1.4)
First- or second-degree	30.3 (141)	23.0 (94)	1.6 (1.2–2.2)

*Adjusted for state, reference age, country of birth, education, marital status, age at menarche, number of live births, breast lump—benign, and height [*see (4)*].

As well as family history, other factors were found to be predictive of case–control status, but their effects were clearer when modeled together in a multiple logistic regression, fitting linear effects across ordered categories of exposure rather than different risk estimates for each category *(4)*.

The odds ratios and relative risk estimates for all those risk factors considered to make up the parsimonious model are shown in Table 3. Having ever had a child, on its own, appeared to be weakly predictive and of marginal statistical significance *(4)*. Once the number of live births was taken into account, however, it was seen that the first birth was associated with a 65% increase in risk but that each subsequent birth reduced the risk by 23% [*see* Fig. 1 in *(4)*]. Current use of oral contraceptives was associated with an increase in risk of 36%, but this finding was of marginal significance.

Height was associated with an increased risk of about 3% per cm [*see* Fig. 2A in *(4)*]. There was limited evidence of a decreased risk with weight or body mass index (BMI), as had previously been reported in the literature. Fitting height and weight (or height and BMI) together, however, enhanced the height effect and suggested a protective effect of weight limited to women in the highest quintile [*see* Fig. 2B in *(4)*].

Table 3. Estimates of log odds ratios and relative risk for the risk factors included in the parsimonious model for the Australian Breast Cancer Family Study, 1992–1995

Risk factor	Log odds ratio* (standard error)	Relative risk* (95% confidence interval)
Ever had live birth	0.50 (0.22)	1.65 (1.06–2.56)
Number of children—1	−0.26 (0.10)	0.77 (0.63–0.94)
Current oral contraceptive use	0.31 (0.17)	1.36 (0.97–1.92)
Height, cm	0.027 (0.011)	1.03 (1.005–1.050)
Affected first-degree relative	1.05 (0.27)	2.86 (1.67–4.90)

*Adjusted for state, reference age, country of birth, education, and marital status [*see (4)*].

We were aided in these analyses and interpretations by use of floating-point standard errors *(18)* that present indications of uncertainty of risk estimates for each category while taking into account adjustments for other risk factors; *see* Figs. 2A and 2B in *(4)*. We calculated approximate standard errors with the use of a simple formula based on an heuristic argument (details available from the first author on request).

We were motivated in our "linear modeling of effects" in part by parsimony and in part by our future intentions to examine whether the effects of these epidemiologic risk factors depend on, or are "modified by," genetic factors. The power to detect such interactions or modifications of risk will be optimized if the nongenetic effect can be well summarized by just one parameter.

Some Findings to Date: Genetic

Some genetic analyses have now been undertaken in the DNA extracted from blood samples collected by the ABCFS between 1992 and 1995. For example, it had been suggested that the codon 325 polymorphism in the estrogen receptor (ER) gene was predictive of breast cancer in women with a family history of breast cancer *(19)*. That observation was based on a series of affected subjects only, selected from a clinic and for whom family cancer history was based on self-report and not verified. Our much larger study *(11)*, using population-based sampling and trying to validate family history going back two generations, found no supporting evidence despite having sufficient statistical power to detect effects one-half the magnitude observed in the hypothesis-generating study. We also found no evidence for an effect of the androgen receptor exon 1 CAG repeat length on risk of breast cancer *(20)*. Finally, using new methods to size minisatellite alleles, we found no evidence to support an effect of the rare alleles at the HRAS1 locus on the risk of early onset breast cancer, despite having ample statistical power to detect an effect of the size predicted by a meta-analysis of previous studies of breast cancer with the use of a less sensitive technique to classify alleles *(21)*.

We have also been conducting extensive mutation testing in these population-based samples *(16,22)*. We estimated the average risk of breast cancer in women who carry a protein-truncating mutation in BRCA1 or BRCA2, the supposedly "high-risk genes for familial breast cancer," with the use of new statistical approaches *(16,17)*. Our estimate, based on studying the relatives of those cases with breast cancer diagnosed before the age of 40 years who were found by us to carry a germline protein-truncating mutation, was about one half that derived from analyses of data from atypical multiple-case families *(17)*.

The striking finding of our study is that the so-called "high-risk" families that contain multiple cases of breast cancer are *not* typical of the families in the general population in which BRCA1 or BRCA2 mutations are segregating. More often than not, the mutation-carrying case diagnosed before the age of 40 years does not have a family history of breast cancer, even going back two generations. It will be of interest to see if this finding holds for case subjects with later age at diagnosis.

Fig. 1 illustrates schematically the intermixing of "familial," "sporadic," and "hereditary" cases of breast cancer in the population, based on our mutation testing. "Familial" cases are those with an affected first- or second-degree relative at the time of diagnosis. "Sporadic" cases are those with no such known family history. As shown in Table 4, the sporadic cases comprise about 70% of all women under the age of 40 years at diagnosis *(4)*, and about one in 20 of these cases carry a BRCA1 or BRCA2 mutation, i.e., are "hereditary." In total, there are more hereditary cases among the sporadic group than among the familial group! In about 28% of all early-onset cases, only one relative is affected. About one in 16 of these familial cases

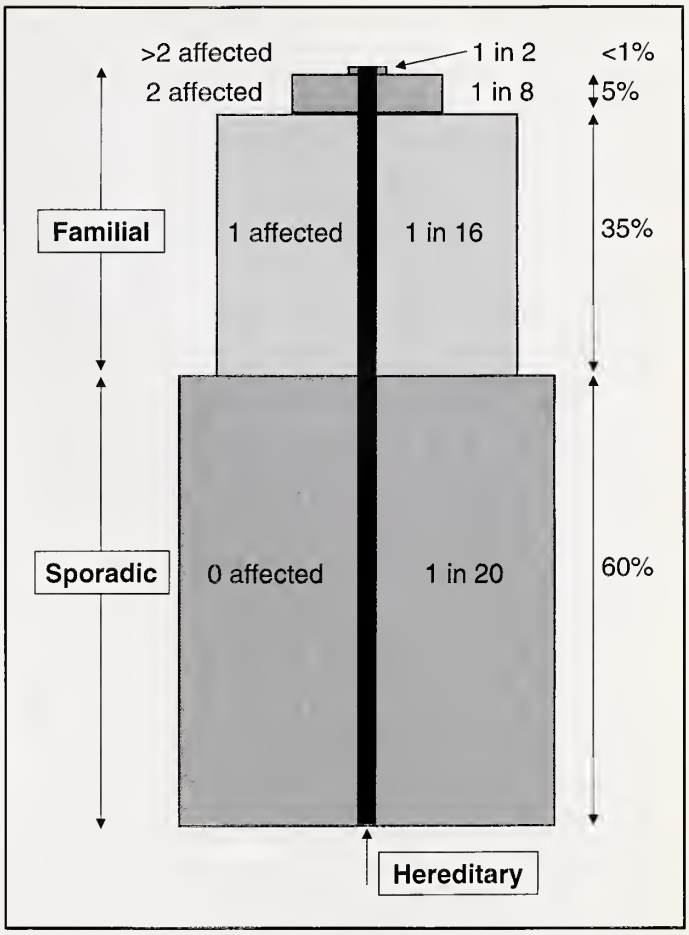


Fig. 1. Women with breast cancer diagnosed before the age of 40 years, categorized according to number of affected first- or second-degree female relatives (0, 1, 2, or >2). The relative areas of the **boxes** indicate the proportion of all affected cases in each category of family history. "Familial" cases have one or more affected first- or second-degree relatives and make up about 30% of all cases *(4)*. "Sporadic" cases have no family history. "Hereditary" cases have a germline mutation in either BRCA1 or BRCA2 and can occur in any category. Although they are more likely to occur if the case is familial, in absolute numbers, most hereditary cases are sporadic. The proportion of hereditary cases in each category of family history, in terms of 1 in *x*, is shown in the **right-hand half of each box.** The proportion of all hereditary cases for each category of family history, in terms of a percentage, is shown on the **right-hand side of the figure.**

Table 4. Approximate breakdown of mutation status in BRCA1 or BRCA2 for women with breast cancer diagnosed before age 40 years, by family history, based on the Australian Breast Cancer Family Study, 1992–1995

No. of affected first- or second-degree female relatives	Proportion of all cases, %	Proportion who carry a mutation in BRCA1/2	Proportion of all mutation carriers, %
0	70	~1 in 20	60
1	28	~1 in 16	35
2	2	~1 in 8	5
>2*	<1	~1 in 2	<1

*Based on the Breast Cancer Linkage Consortium [*see (17)*].

carries a mutation, the more so if the other affected family member is a first-degree relative (one in 10) than a second-degree relative (one in 20). If two relatives are affected, then about one in eight carries a mutation, but the latter group constitutes only 2% of all cases. Cases with three or more affected relatives make up less than 1% of all cases, even those of early onset, and less than 1% of all mutation-carrying cases. [In Fig. 1, we have presumed that about one half of these carry a mutation in either BRCA1 or BRCA2, based on the findings of the Breast Cancer Linkage Consortium; see *(17)*.] That is, so-called "high-risk families" constitute less than 1% of all BRCA1 or BRCA2 mutation carriers, whereas 60% of mutation carriers are in families with *no* family history.

Fig. 1 resembles the shape of a three-story tower with a broad base representing the sporadic cases and a small top section representing those with the strongest family history. It will be of interest to find out what the shape of this tower looks like for case patients with a later age at diagnosis as well as for other cancers and other genes. Does the tower become fatter or thinner, do the familial components get bigger or smaller, and does the heritable core get wider or narrower?

Our findings have implications for genetic testing and counseling, raising the possibility that risk prediction in mutation carriers may need to vary according to previous family history of breast or other cancers. Clearly, systematic prospective studies of multiple-case mutation-carrying families are needed to assess if they really deserve to be classified as "high risk." It also raises the interesting prospects that cancer risk may vary considerably between mutations, even between those that are protein-truncating, or that genetic and environmental risk factors may exist that modify risk in mutation carriers. Extensive mutation testing in large population-based family studies such as ours, with information on both genetic and questionnaire-derived risk factors, will be important in clarifying these issues.

Conclusion

Population-based multicenter, multidisciplinary enterprises, such as those being collected by the CFRBCS, may become a model for the future research in cancer epidemiology and cancer genetics. Cancer epidemiology will benefit by moving from considering the individual to the family as the unit of interest and, therefore, allowing assessment of the role of genetic factors, especially if blood or tissue samples, or both, are available. Cancer genetics will benefit immeasurably by the fact that findings can be immediately put into a population and clinical perspective.

References

(1) Morton NE. Outline of genetic epidemiology. Basel (Switzerland): Karger; 1982.

(2) Last J. A dictionary of epidemiology. Oxford (UK): Oxford University Press; 1983.

(3) Hopper JL, Giles GG, McCredie MR, Boyle P. Background, rationale and protocol for a case–control–family study of breast cancer. Breast 1994;3: 79–86.

(4) McCredie MR, Dite GS, Giles GG, Hopper JL. Breast cancer in Australian women under the age of 40. Cancer Causes Controls 1998;9:189–98.

(5) Bonney GE. Regressive logistic models for familial disease and other binary traits. Biometrics 1986;42:611–25.

(6) Hopper JL, Jenkins MA, Macaskill GT, Giles GG. Regressive logistic modelling of familial aggregation for smoking in a population-based sample of nuclear families. J Epidemiol Biostats 1997;2;45–52.

(7) Hopper JL, Carlin JB. Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. Am J Epidemiol 1992;136:1138–47.

(8) Whittemore AS, Halpern J. Multi-stage sampling in genetic epidemiology. Stat Med 1997;16:153–67.

(9) Siegmund KD, Whittemore AS, Thomas DC. Multistage sampling for disease family registries. Monogr Natl Cancer Inst 1999;26:43–8.

(10) Aitken JF, Youl P, Green A, MacLennan R, Martin NG. Accuracy of case-reported family history of melanoma in Queensland, Australia. Melanoma Res 1996;6:313–7.

(11) Southey MC, Batten LE, McCredie MR, Giles GG, Dite G, Hopper JL, et al. Estrogen receptor polymorphism at codon 325 and risk of breast cancer in women before age forty. J Natl Cancer Inst 1998;90:532–6.

(12) Thompson EA. Pedigree analysis in human genetics. Baltimore (MD): The Johns Hopkins University Press; 1986.

(13) Armes JE, Egan AJ, Southey MC, Dite GS, McCredie MR, Giles GG, et al. The histologic phenotypes of breast carcinoma occurring before age 40 years in women with and without BRCA1 or BRCA2 germline mutations: a population-based study. Cancer 1998;83:2335–45.

(14) Armes JE, Trute L, White D, Southey MC, Hammet F, Tesoriero A. et al. Distinct molecular pathogeneses of early-onset breast cancers in BRCA1 and BRCA2 mutation carriers: a population-based study. Cancer Res 1999; 59:2011–7.

(15) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Monogr Natl Cancer Inst 1999;26:31–7.

(16) Hopper JL, Southey MC, Dite GS, Jolley DJ, Giles GG, McCredie MR et al. for the Australian Breast Cancer Family Study. Population-based estimate of the average age-specific cumulative risk of breast cancer for a defined set of protein-truncating mutations in BRCA1 and BRCA2. Cancer Epidemiol Biomark Prevent 1999;8:741–47.

(17) Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, et al. and the Breast Cancer Linkage Consortium. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. Am J Hum Genet 1998;62:676–89.

(18) Easton DF, Peto J, Babiker AG. Floating absolute risk: an alternative to relative risk in survival and case–control analysis avoiding an arbitrary reference. Stat Med 1991;10:1025–35.

(19) Roodi N, Bailey LR, Kao WY, Verrier CS, Yee CJ, Dupont WD, et al. Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. J Natl Cancer Inst 1995;87:446–51.

(20) Spurdle AB, Dite GS, Chen X, Mayne CJ, Southey MC, Batten LE, et al. Androgen receptor exon 1 CAG repeat length is not a risk factor for breast cancer in women under the age of 40. J Natl Cancer Inst 1999;91:961–6.

(21) Firgaira FA, Seshadri R, McEvoy CR, Dite GS, Giles GG, McCredie MR, et al. HRAS1 rare minisatellite alleles and breast cancer in Australian women under age forty years. J Natl Cancer Inst 1999;24:2107–11.

(22) Southey MC, Tesoriero AA, Andersen CR, Jennings KM, Brown SM, Dite GS, et al. BRCA1 mutations and other sequence variants in a population-based sample of Australian women with breast cancer. Br J Cancer 1999; 79:34–9.

Combined Influence of Genetic and Dietary Factors on Colorectal Cancer Incidence in Japanese Americans

Loïc Le Marchand

Colorectal cancer (CRC) rates for Japanese migrants to the United States increased rapidly to surpass the level of the host population. CRC rates for the Japanese in Hawaii and California are now the highest in the world. Rates for this disease have also increased in Japan, presumably as the result of the westernization of the diet. A series of population-based studies in Hawaii was undertaken to determine which dietary factors are responsible for this remarkable susceptibility of the Japanese to CRC and whether genetic factors are also involved. A first case–control study suggested that a high intake of red meat is a major risk factor for the disease in Hawaii Japanese men and that family history of CRC among first-degree relatives may strongly modify this association. A case–control family study is currently being completed to explore further the interaction between family history and the intake of red meat after adjustment for environmental covariates among family members. Also, a segregation analysis will guide gene discovery studies among high-risk Japanese families being recruited in the Cooperative Family Registry for Colorectal Cancer Studies. Retrospective and prospective studies are also ongoing to test associations of CRC with specific polymorphisms in genes controlling the metabolic activation or detoxification of the carcinogens associated with a diet high in red meat. Preliminary results suggest an association of the combined rapid NAT2 genotype and rapid CYP1A2 phenotype with CRC in individuals consuming well-done red meat. Populations in which dramatic changes in cancer incidence have occurred may offer opportunities to identify gene–environment interactions. [Monogr Natl Cancer Inst 1999;26:101–5]

Gene–environment interactions have rarely been studied as possible explanations for epidemiologic patterns of cancer in populations. More commonly, investigators have focused on environmental factors in attempting to explain temporal trends or international differences in cancer rates. This approach has been useful in identifying reproductive, lifestyle, and infectious risk factors for cancer and has led to the acceptance of the predominant role of the environment in cancer causation. However, genetic factors may play an important role in modifying the cancer risk of a population. This role is epitomized by the skin cancer epidemic that has resulted from an increase in recreational UV exposure among Western populations during the past 30 years. This increase in rates has been largely limited to populations with light pigmentation, a genetically determined trait.

The public health impact of a gene–environment interaction may be large if the at-risk genotype and exposure are common. Thus, such interactions could be especially relevant to populations in which changes in cancer rates have been particularly dramatic. The rapid increase in colorectal cancer (CRC) rates experienced by Japanese migrants to Hawaii might offer a unique opportunity to investigate such interactions.

Japanese Migrants to Hawaii

Between 1886 and 1924, approximately 214 000 rural Japanese were brought to Hawaii from the southern prefectures of Hiroshima, Fukuoka, Yamaguchi, and Okinawa to provide laborers to the sugar plantations *(1)*. No sizable migration from Japan to Hawaii has taken place since the 1920s and, up until the most recent generation, there has been very little admixture of the Japanese with other ethnic groups in Hawaii. Thus, these migrants and their descendants, at least in the age group affected by common cancers, appear to have been quite homogeneous genetically. Because of a high birth rate during the first half of the century, Japanese now represent 30% of the state population.

Descriptive Epidemiology of CRC in Hawaii Japanese

The incidence of CRC for Japanese in Hawaii increased markedly on migration, presumably, as a result of a dietary change *(2)*. However, whereas Japanese incidence rates for other cancer sites, such as breast and stomach cancers, took two generations or more to approach those of the Caucasians, Japanese CRC rates increased faster and matched those of Caucasians as early as the first generation (Fig. 1) *(2)*. The same observations were reported for Japanese migrants to California *(3,4)*. Japanese Americans in Hawaii and Los Angeles in 1983–1987 had the first and second highest incidence rates for CRC, respectively, among more than 175 populations followed for cancer incidence worldwide *(5)*. Similarly, as the diet in Japan has become more westernized, incidence for this cancer in Japan over the past 20 years has increased even more rapidly than for other cancers *(6,7)*. As shown in Fig. 2, in comparison to several representative populations, CRC incidence in Japan has now reached the level of the U.K. *(5)*. These observations suggest that Japanese may be particularly susceptible to CRC because of an especially pronounced exposure to lifestyle risk factors in the United States, to the modifying influence of genetic susceptibility factors, or to both *(8)*. A strong genetic predisposition as a result of a founder effect in some large families of Japanese migrants to Hawaii could also be present but seems less likely because an increase in CRC rates has taken place in Japanese populations elsewhere.

Dietary surveys conducted during the 1970s and 1980s have consistently shown that the first and subsequent generations of Japanese migrants to Hawaii have retained some elements of their ethnic diet, and, in particular, their fat intake has remained

Correspondence to: Loïc Le Marchand, M.D., Ph.D., Etiology Program, Cancer Research Center of Hawaii, University of Hawaii, 1236 Lauhala St., Suite 407, Honolulu, HI 96813 (e-mail: loic@crch.hawaii.edu).

See "Note" following "References."

© Oxford University Press

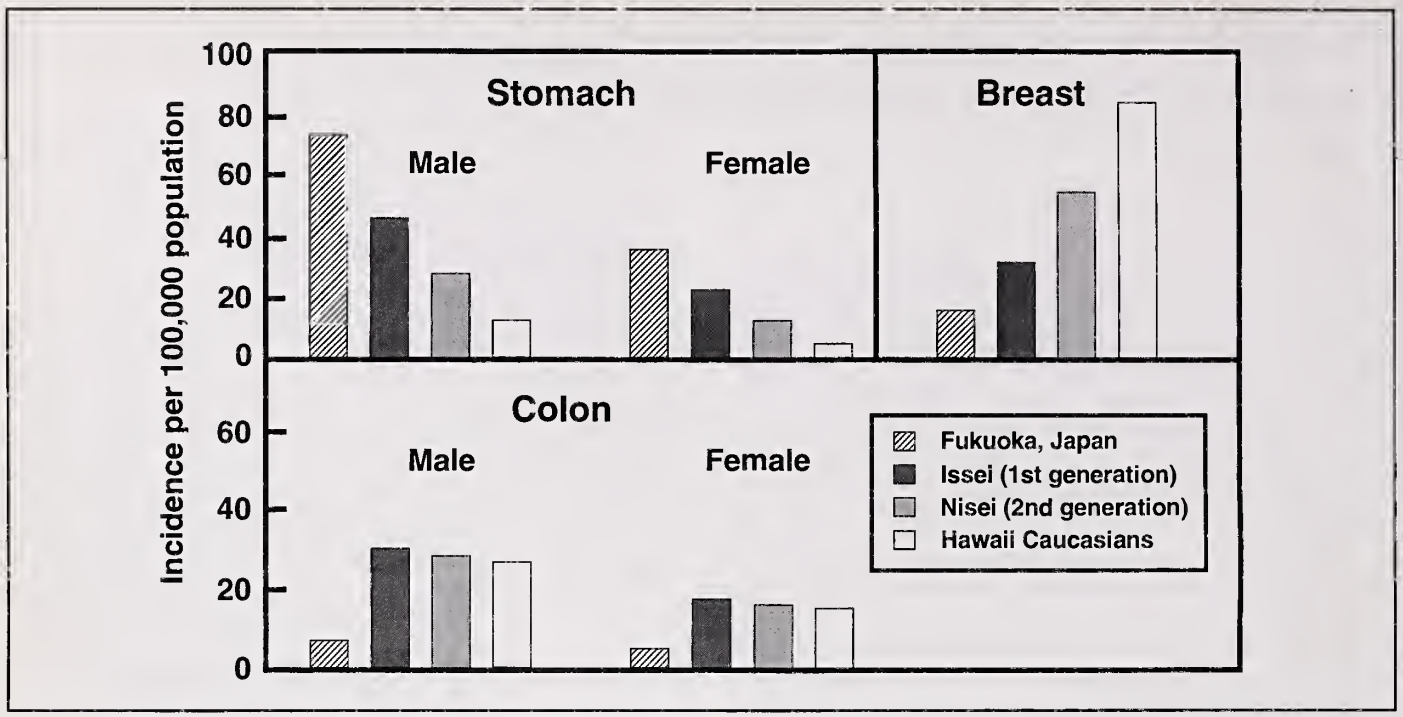


Fig. 1. Age-adjusted incidence rates for Japanese migrants to Hawaii and their source and host populations for the period 1973–1977 [adapted from *(2)*].

distinctly lower than that of Caucasians *(1,9)*. With the objective of identifying the specific lifestyle changes responsible for the sharp increase in CRC risk in Japanese Americans, we first conducted a population-based, case–control study of the disease in Hawaii.

Case–Control Study of Lifestyle and CRC

We interviewed 698 men and 494 women diagnosed from 1987 through 1991 with adenocarcinoma of the colon or rectum and 1192 population control subjects matched to case patients on

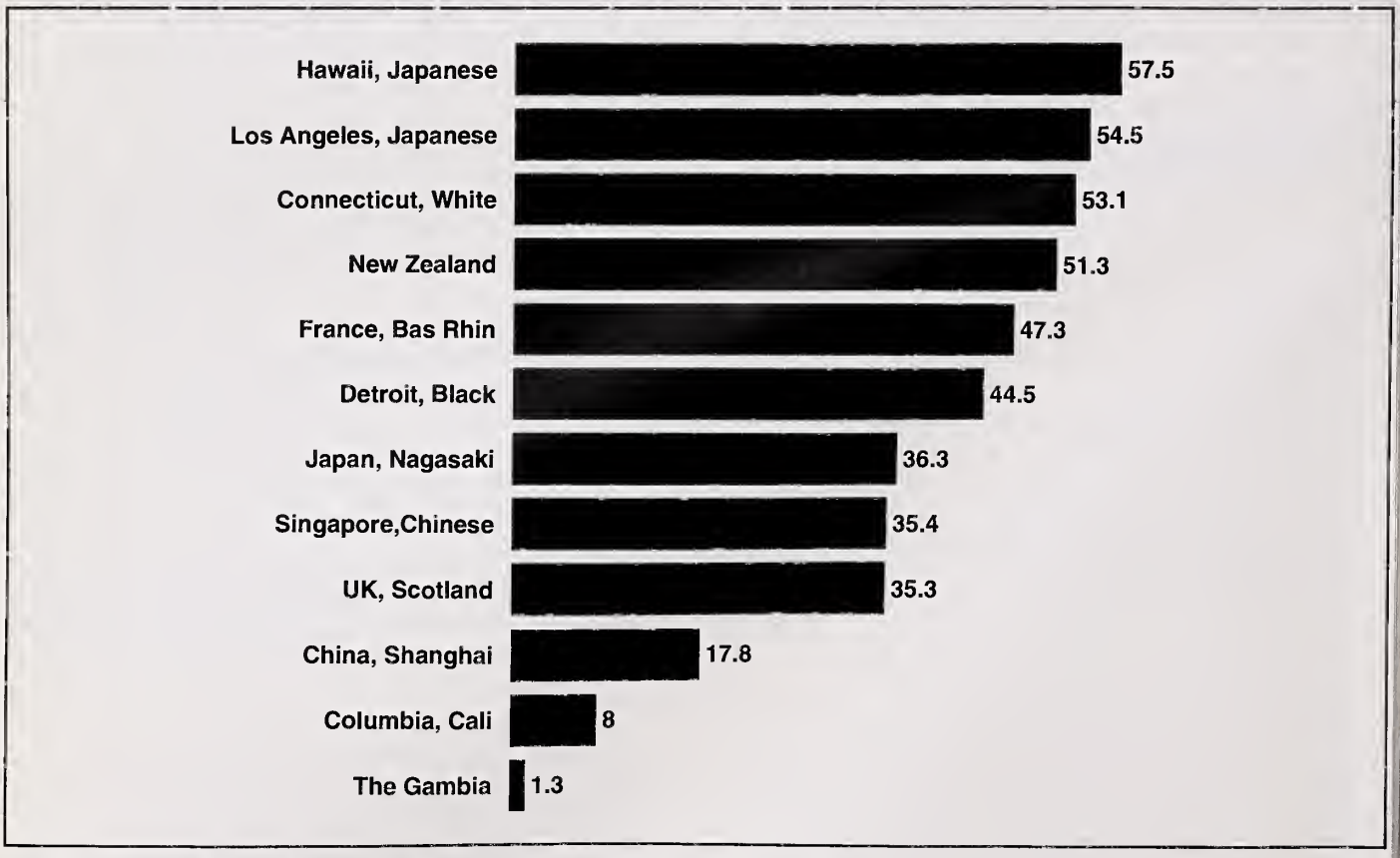


Fig. 2. Male colorectal cancer incidence in selected countries for the period from 1983 through 1987 *(5)*.

age, sex, and ethnicity. Fifty percent of the subjects were of Japanese ancestry, whereas 25% were Caucasians and the rest Filipino, Hawaiians, or Chinese. The questionnaire inquired about usual intake of more than 280 food items as well as about detailed ethnic background, lifetime recreational and occupational physical activity, smoking, medical history, family history of cancer, and body size. Consistent with studies conducted in other populations *(10)*, the main lifestyle factors found to be directly associated with risk in this study were the intake of beef and processed meat (in males only), consumption of eggs, smoking, alcohol use, and obesity *(11–13)*. Inverse associations were found with the intake of vegetable fiber and calcium and with lifetime recreational physical activity *(12–14)*. With the use of the population controls to compare the dietary intakes of the various ethnic groups, we found, as expected, that Japanese had a lower fat intake than Caucasians (Table 1). More surprising was the finding that Japanese consumed more red meat, particularly beef, and processed meats as well as less fiber and vegetables than Caucasians (Table 1). Thus, these dietary factors could contribute to the higher CRC risk of the Japanese.

For all subjects combined, we also found a 2.5-fold increased risk (95% confidence interval [CI] = 1.2–2.9) of CRC in first-degree relatives of case patients compared with relatives of control subjects *(15)*. The magnitude of this association was greater for case patients with a right-sided tumor, multiple tumors, and, especially, early age at diagnosis. Of interest, we also found that the risk associated with family history was stronger in Japanese than in Caucasians, with an odds ratio of 3.0 (95% CI = 2.1–4.6) and 1.8 (95% CI = 1.2–2.9), respectively *(15)*. The difference between these odds ratios was of borderline statistical significance (P = .07). These results suggest that the genetic component of the disease may be stronger in Japanese than in Caucasians, perhaps because of a gene–environment interaction.

A recent analysis *(14)* also explored interactions between the lifestyle risk factors identified in this study and the family history of the disease among first-degree relatives. In an analysis conducted on all ethnic groups and adjusting for all lifestyle main effects and interactions, statistically significant interactions with family history were observed for the intakes of beef and alcohol in males. Only the data on beef will be discussed here. Relative to men without a family history of the disease and whose intake of beef fell in the lower third, odds ratio for CRC for men with a family history and in the upper tertile of intake was 10.8 (95% CI = 4.2–27.6) for beef intake (Fig. 3). The corresponding odds ratio for men without a family history and in the upper tertile of beef intake was 1.5 (1.0–2.3). This interaction appeared to be stronger in Japanese than in Caucasians, although the power of these ethnic-specific analyses was limited. No interactions were found in women, possibly because of the smaller sample size and lower consumption of beef or because of some underlying biologic differences *(14)*. This analysis suggested that family history increases risk of CRC in men mainly through a strong interaction with a high intake of beef and alcohol *(14)*. However, because of the high number of statistical tests performed and the discrepancy between sexes, these findings need replication.

In summary, this first case–control study suggested that a high intake of red meat may contribute to the high CRC risk of Japanese in Hawaii and that family history of the disease among first-degree relatives may strongly modify this association. To investigate these associations further, we are currently conducting two types of studies: a case–control family study is exploring the familial aggregation of CRC and interaction of family history with lifestyle factors while taking into account the potential confounding effects of covariates in relatives, and a second case–control study and a prospective study are testing associations with candidate genes that may modify the association between the intake of red meat and CRC.

Case–Control Family Study

One objective of this genetic epidemiology study is to separate the environmental (dietary and lifestyle) and genetic components of the familial aggregation of CRC and explore gene–environment interactions for the disease in our population. To increase the efficiency of the study, only case patients diagnosed before age 60 years and their sex-, age- and race-matched control subjects are included as probands. Subjects are of Japanese, Caucasian, Hawaiian, or Chinese origin. Parents and siblings are contacted by mail to collect information on their demographics, medical background, family history of cancer, and environmental exposures (diet, smoking, physical activity, and body size). The statistical analysis will explore patterns of familial aggregation after adjusting for environmental covariates for both probands and relatives and will investigate interactions between family history and lifestyle factors, such as the intake of red meat *(16,17)*. The second objective of this study is to perform a segregation analysis after adjusting for environmental covariates to explore various models of inheritance for susceptibility to CRC in our population *(18)*. This analysis may guide gene dis-

Table 1. Geometric means* for daily intake of Japanese and Caucasians in Hawaii

	Male			Female		
	Japanese (n = 347)	Caucasian (n = 180)	P	Japanese (n = 244)	Caucasian (n = 116)	P
Calories, Kcal	2101	2248	.04	1580	1743	.02
Total fat, g	77	86	<.001	58	68	<.001
Saturated fat, g	21	25	<.001	15	20	<.001
Red meat, g	46	40	.10	31	27	.08
Beef, g	12	6	<.001	8	3	<.001
Processed meats, g	23	17	.001	12	9	.02
Dietary fiber, g	16	19	<.001	15	16	.11
Vegetables, g	184	217	<.001	183	211	.01
Fruits, g	222	251	.13	260	274	.51

*Based on population controls and adjusted for age by covariance analysis and by calories (except total calories) by the method of residuals.

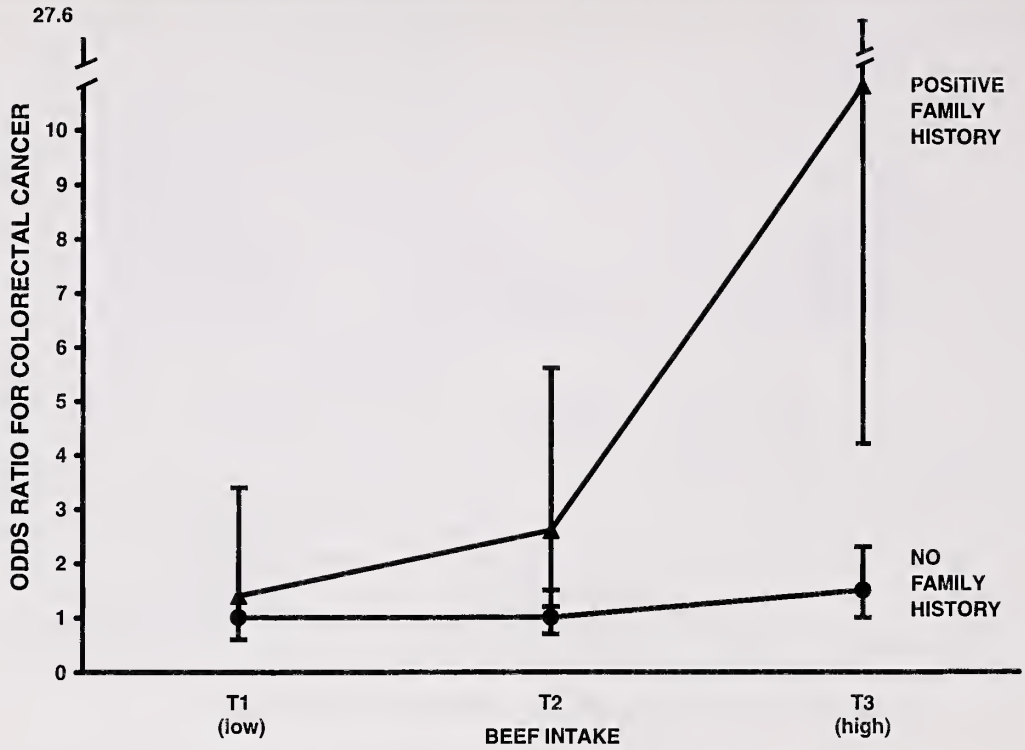


Fig. 3. Odds ratios for colorectal cancer by tertile of beef intake and first-degree family history of colorectal cancer among men in Hawaii *(15)*.

covery studies among the high-risk Hawaii families that are being enrolled in the Cooperative Family Registry for Colorectal Cancer Studies (CFRCCS) supported by the National Cancer Institute.

Retrospective and Prospective Association Studies

These studies are testing whether exposure to a diet high in red meat or well-done meat, combined with specific genotypes associated with altered activities of type I and II carcinogen metabolizing enzymes, may explain the high CRC risk of Japanese Americans. The well-established association of red meat with CRC *(10)* may be due in part to the formation of heterocyclic aromatic amines (HAAs) by pyrolysis of creatin and creatinine in meats cooked at high temperature (e.g., panfrying, broiling, or barbecuing) *(19)*. Levels of HAAs in meat greatly depend on the duration of cooking. HAAs are highly carcinogenic to the large bowel in rodents and are suspected to be carcinogenic to humans as well *(19)*. HAAs require metabolic activation by CYP1A2 and *N*-acetyltransferase to bind to DNA and exert their genotoxicity *(20)*. It has been hypothesized that persons with the rapid N-oxidation and rapid NAT2 phenotypes are at higher risk for CRC, especially when they consume their meat well done *(21,22)*. Of interest, the fast acetylator genotype is especially common in Japanese (90%) compared with Caucasians (45%), whereas the frequency of the CYP1A2 susceptible phenotype is similar in the two groups *(23–25)*.

We are completing a second case–control study of CRC, following the same population-based design and including the same ethnic groups as in the first study. Case patients were diagnosed between 1994 and 1999. An interviewer-administered, quantitative food-frequency questionnaire assesses total diet as well as intake of red meat cooked by different methods to various levels of doneness. Blood and urine samples are collected. Subjects are genotyped for NAT2*4, NAT2*5A, NAT2*6A, and NAT2*7A by a polymerase chain reaction-based method *(26)* and phenotyped for CYP1A2 activity by the caffeine test *(27)*. Preliminary results from this new study appear to confirm an association of the combined rapid NAT2 genotype and rapid CYP1A2 phenotype with CRC that is limited to individuals consuming their red meat well done (Table 2). The prevalence of eating well-done meat was similar in Japanese and Caucasians, and the "meat preference" of those who reported not to eat red meat was set to "rare/medium" in this analysis. No main effect was found for the NAT2 rapid acetylator genotype alone. Future analyzes will estimate the CRC risk associated with intakes of the main HAAs (PhIP and MeIQx).

Table 2. Odds ratios* (95% confidence interval) for colorectal cancer by NAT2 genotype and CYP1A2 phenotype

		Meat preference	
NAT2/CYP1A2	All (150/210)†	Well-done (43/57)	Rare/medium (106/152)
Others	1.0	1.0	1.0
Rapid/rapid‡	1.4 (0.8–2.6)	3.6 (1.1–11.6)	1.4 (0.8–2.4)

*Adjusted for age, sex, and race.
†No. of case patients/No. of control subjects.
‡NAT2*4/*4 genotype and rapid CYP1A2 phenotype (urinary 17X + 17U)/137X > median) *(26,27)*.

Because exposure information in case–control studies may be plagued by differential recall, we are attempting to reproduce this association in a prospective study in which diet information was collected before diagnosis. This study is following for cancer occurrence approximately 215 000 Japanese, Caucasian, African-American, Hispanic, and Native Hawaiian residents of Hawaii and Los Angeles who mailed back a detailed epidemiologic questionnaire between 1993 and 1996 *(28)*. Blood samples are being collected from incident CRC case patients and a subset of the cohort members to conduct a nested case–control study.

Although some of the findings obtained to date are still preliminary, this research suggests that consumption of well-done red meat, combined with a specific genotype (NAT2*4/*4 and rapid CYP1A2), may increase CRC risk substantially. Because both this dietary exposure and genotype are more common in Japanese Americans, they may contribute to the high risk of the disease in this population. Other polymorphic metabolic genes for which the at-risk genotype is common in Japanese may also be involved *(29–31)*. Linkage studies in multiplex Japanese families may also identify other genes that may contribute to the strong familial aggregation of the disease in this ethnic group in Hawaii.

References

(1) Nordyke EC. The peopling of Hawaii. Honolulu (HI): University Press of Hawaii; 1977.

(2) Kolonel LN, Hankin JH, Nomura AM. Multiethnic studies of diet, nutrition and cancer in Hawaii. In: Hayashi Y, Nagao M, Sugimura T, Takayama S, Tomatis L, Wattenberg LW, et al. editors. Diet, nutrition and cancer. Tokyo (Japan): Japan Sci Soc Press; 1986. p. 29–40.

(3) Shimizu H, Mack TM, Ross RK, Henderson BE. Cancer of the gastrointestinal tract among Japanese and white immigrants in Los Angeles County. J Natl Cancer Inst 1987;78:223–8.

(4) Shimizu H, Ross RK, Bernstein L, Yotani R, Henderson BE, Mack TM. Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. Br J Cancer 1991;63:963–6.

(5) Parkin DM, Muir CS, Whelan SL, Gao YT, Ferlay J, Powell J. Cancer incidence in five continents, Vol. VI. Lyon (France): International Agency for Research on Cancer; 1992.

(6) Willett W. The search for the causes of breast and colon cancer. Nature 1989;338:389–94.

(7) Broder S. Perspectives on cancer in Japan and the United States. Jpn J Cancer Res 1993;84:821–30.

(8) Le Marchand L, Kolonel LN. Cancer among Japanese migrants to Hawaii: gene–environment interactions. Rev Epidemiol Santé Publique 1992;40: 425–30.

(9) Kolonel LN. Variability in diet and its relation to risk in ethnic and migrant groups. In: Woodhead AD, Bender MA, Leonard RC, editors. Phenotypic variations in populations. New York (NY): Plenum; 1988.

(10) World Cancer Research Fund/American Institute for Cancer Research. Food, nutrition and the prevention of cancer: a global perspective. Washington (DC): American Institute for Cancer Research; 1997, p. 216–51.

(11) Le Marchand L, Wilkens LR, Hankin JH, Kolonel LN, Lyu LC. A case–control study of diet and colorectal cancer in a multiethnic population in Hawaii (United States): lipids and foods of animal origin. Cancer Causes Control 1997;8:637–48.

(12) Le Marchand L, Hankin JH, Wilkens LR, Kolonel LN, Englyst HN, Lyu LC. Dietary fiber and colorectal cancer risk. Epidemiology 1997;8:658–65.

(13) Le Marchand L, Wilkens LR, Kolonel LN, Hankin JH, Lyu LC. Associations of sedentary lifestyle, obesity, smoking, alcohol use, and diabetes with the risk of colorectal cancer. Cancer Res 1997;57:4787–94.

(14) Le Marchand L, Zhao LP, Quiaoit F, Wilkens LR, Kolonel LN. Family history of colorectal cancer in the multiethnic population of Hawaii. Am J Epidemiol 1996;144:1122–8.

(15) Le Marchand L, Wilkens LR, Hankin JH, Kolonel LN, Lyu LC. Independent and joint effects of family history and lifestyle on colorectal cancer risk: implications for prevention. Cancer Epidemiol Biomarkers Prev 1999; 8:45–51.

(16) Zhao LP, Le Marchand L. An analytical method for assessing patterns of familial aggregation in case–control studies. Genetic Epidemiol 1992;9: 141–54.

(17) Zhao LP, Hsu L, Davidov O, Potter J, Elston R, Prentice RL. Population-based family study designs: an interdisciplinary research framework for genetic epidemiology. Genetic Epidemiol 1997;14:365–88.

(18) Zhao LP, Quiaoit F, Hsu L, Davidov O, Holte S. A population-based family study of a common oligogenic disease (II): the segregation analysis. Genetic Epidemiol 1997;14:945–9.

(19) Sugimura T. Overview of carcinogenic heterocyclic amines. Mutation Res 1997;376:211–9.

(20) Minchin RF, Reeves PT, Teitel CH, McManus ME, Mojarrabi B, Ilett KF, et al. *N*- and *O*-acetylation of aromatic and heterocyclic amine carcinogens by human monomorphic and polymorphic acetyltransferases expressed in cos-1 cells. Biochem Biophys Res Comm 1992;185:839–44.

(21) Ilett KF, David BM, Dechton P, Castleden WM, Kwa R. Acetylation phenotype in colorectal carcinoma. Cancer Res 1987;47:1466–9.

(22) Lang NP, Butler MA, Massengill J, Lawson M, Stotts RC, Hauer-Jensen M, et al. Rapid metabolic phenotypes for acetyltransferase activity and cytochrome P4501A2 and putative exposure to food-borne heterocyclic amine increase the risk of colorectal cancer or polyps. Cancer Epidemiol Biomarkers Prev 1994;3:675–82.

(23) Lin HJ, Han CY, Lin BK, Hardy S. Ethnic distribution of slow acetylator mutations in the polymorphic *N*-acetyltransferase (NAT2) gene. Pharmacogenetics 1994;4:125–34.

(24) Le Marchand L, Sivaraman L, Franke AA, Custer LJ, Wilkens LR, Lau AF, et al. Predictors of *N*-acetyltransferase activity: should caffeine phenotyping and NAT2 genotyping be used interchangeably in epidemiological studies? Cancer Epidemiol Biomarkers Prev 1996;5:449–55.

(25) Le Marchand L, Franke A, Custer L, Wilkens LN, Cooney RV. Lifestyle and nutritional correlates of cytochrome CYP1A2 activity: inverse associations with plasma lutein and alpha-tocopherol. Pharmacogenetics 1997; 7:11–9.

(26) Bell DA, Taylor JA, Buttler MA, Stephens EA, Wiest J, Brubaker LH, et al. Genotype/phenotype discordance for human arylamine *N*-acetyltransferase (NAT2) reveals a new slow-acetylator allele common in African-Americans. Carcinogenesis 1993;14:1689–92.

(27) Butler MA, Lang NP, Young JF, Caporaso NE, Vineis P, Hayes RB, et al. Determination of CYP1A2 and NAT2 phenotypes in human populations by analysis of caffeine urinary metabolites. Pharmacogenetics 1992;2:116–27.

(28) Kolonel LN, Henderson BE, Hankin JH, Nomura AMY, Wilkens LR, Pike MC, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol. In press 1999.

(29) Bell DA, Stephens EA, Castranio T, Umbach DM, Watson M, Deakin M, et al. Polyadenylation polymorphism in the acetyltransferase 1 gene (NAT1) increases risk of colorectal cancer. Cancer Res 1995;55:3537–42.

(30) Sivaraman L, Leatham MP, Yee J, Wilkens LR, Lau AF, Le Marchand L. CYP1A1 genetic polymorphisms and *in-situ* colorectal cancer. Cancer Res 1994;54:3692–5.

(31) Le Marchand L, Sivaraman L, Pierce LM, Seifried A, Lum A, Franke AA, et al. Gene–diet interactions and the high colorectal cancer risk of Japanese Americans. In: familial cancer and prevention—molecular epidemiology: a new strategy toward cancer control. New York (NY): John Wiley & Sons; in press 1999.

Note

Supported in part by Public Health Service (PHS) grants P01CA33619, R01CA60987, and U01CA74806 and by PHS contract N01PC67001 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

NIH LIBRARY
3 1496 00834 6895

2746 2304 18
07•12•02 MAB

THE NIH LIBRARY
NATIONAL INSTITUTES OF HEALTH

LIBRARY
NIH
Amazing Research.
Amazing Help.

http://nihlibrary.nih.gov

10 Center Drive
Bethesda, MD 20892-1150
301-496-1080

NIH LIBRARY
3 1496 00915 8596